Received: 22 November 2018 | Revised: 27 January 2020 | Accepted: 3 February 2020

DOI: 10.1002/jgt.22551

ARTICLE

WILEY

Random walk hitting times and effective resistance in sparsely connected Erdős-Rényi random graphs

John Sylvester[1,2]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, United Kingdom

[2]Mathematics Institute, University of Warwick, Coventry, United Kingdom

Correspondence
John Sylvester, Department of Computer Science and Technology, University of Cambridge, 15 JJ Thomson Avenue, Cambridge, CB3 0FD, United Kingdom.
Email: john.sylvester@cl.cam.ac.uk

Funding information
EPSRC, Grant/Award Number: EP/HO23364/1; ERC, Grant/Award Numbers: 639046 (RGGC), 679660 (DYNAMIC MARCH)

Abstract

We prove a bound on the effective resistance $R(x, y)$ between two vertices x, y of a connected graph which contains a suitably well-connected subgraph. We apply this bound to the Erdős-Rényi random graph $\mathcal{G}(n, p)$ with $np = \Omega(\log n)$, proving that $R(x, y)$ concentrates around $1/d(x) + 1/d(y)$, that is, the sum of reciprocal degrees. We also prove expectation and concentration results for the random walk hitting times, Kirchoff index, cover cost, and the random target time (Kemeny's constant) on $\mathcal{G}(n, p)$ in the sparsely connected regime $\log n + \log\log\log n \leq np < n^{1/10}$.

KEYWORDS

effective resistance, hitting time, kirchoff index, random graph, random walk

JEL CLASSIFICATION

05C40; 05C80; 05C81; 60C05; 60J45; 60J85

1 | INTRODUCTION AND RESULTS

The effective resistance $R(x, y)$ between two vertices x, y of a graph $G = (V, E)$ is the energy dissipated by a unit current flow from x to y when all edges have unit resistances. That is,

$$R(x, y) = \inf_{\theta}\left\{\sum_{e\in E}\theta(e)^2 : \theta \text{ is a unit flow from } x \text{ to } y\right\}, \tag{1}$$

see Section 1.3 for a complete mathematical formulation. The effective resistance has connections to Markov chain theory, in particular for infinite graphs the transience or recurrence of a random walk is determined by the resistance from the origin to cut sets at arbitrary distance from the origin [13]. In finite graphs the resistances determine hitting times of random walks [26] and are related to the eigenvalues of the Laplacian [23].

We prove a new bound on effective resistance for graphs G containing a subgraph H with good connectivity properties. The result, Theorem 2.3, may be stated (very) informally as

$$R(x, y) \leq \frac{1}{d(x)} + \frac{1}{d(y)} + \mathrm{Error}_{x,y}(G, H),$$

where $d(\cdot)$ is the degrees of a vertex and $\mathrm{Error}_{x,y}(G, H)$ is an error term. Depending on the graph G and the subgraph H chosen the error may be insignificant compared to the other terms. In that case our bound is essentially tight as $1/(d(x) + 1) + 1/(d(y) + 1)$ is always a lower bound on $R(x, y)$, see Section 2 for a full statement of Theorem 2.3. Although this bound holds for any connected graph, the $\mathrm{Error}_{x,y}(G, H)$ term may dominate; our bound works well for graphs with strong expansion properties. We apply this bound to the random graph $\mathcal{G} \sim \mathcal{G}(n, p)$, that is the simple graph $\mathcal{G}$ on n vertices with law $\mathcal{G}(n, p)$ given by sampling each edge independently with probability p. The random graph $\mathcal{G}(n, p)$ has been extensively studied [4,16,19] and so it is a natural question to determine the effective resistance for such a fundamental graph distribution. We remark that throughout all log's are base e and we define $\varepsilon_n := \varepsilon_n(n, p)$ to be the function

$$\varepsilon_n := \frac{\log n}{np\, \log(np)}. \tag{2}$$

Theorem 1.1. *For any $c > 0$ let $\mathcal{G} \sim \mathcal{G}(n, p)$ with $c \log n \leq np \leq n^{1/10}$. Then for a fixed $i, j \in V$, where $i \neq j$,*

$$\mathbb{P}\left(\left| R(i,j) - \left(\frac{1}{d(i)} + \frac{1}{d(j)}\right)\right| > 9\frac{2 + \varepsilon_n d(i)}{d(i)^2} + 9\frac{2 + \varepsilon_n d(j)}{d(j)^2}\right) = o\left(e^{-np/4}\right) + o\left(n^{-3}\right).$$

Notice if $np = \Omega(\log n)$ then $\varepsilon_n = o(1)$. Theorem 1.1 shows that with high probability (w.h.p.) the main contribution to the effective resistance $R(i, j)$ between vertices $i, j \in V$ comes from the flow through edges connecting i and j to their immediate neighbours.

From the definition (1) of $R(x, y)$ one observes that the contribution to the resistance from each edge in the graph is quadratic in the amount of flow passing through that edge. The main idea of Theorem 2.3 is to show that if a graph contains a subgraph from a certain family well connected graphs then there are many paths between the neighbours of x and y which become edge disjoint away from x and y. If this is the case then the flow can be divided up evenly between the edges close to x and y, further away we use the edge disjoint paths to carry the flow. In a graph with good expansion there should be many paths such paths and so the flow through these edges should be negligible compared to that through edges close to x and y. However since we balanced the flow evenly over edges close to x and y the contribution to $R(x, y)$ from these edges is close to optimal and matches a simple lower bound up to lower order terms. Aside from $\mathcal{G}(n, p)$ our bound on resistance, Theorem 2.3, may potentially be applied to other random graph models such as binomial random intersection graphs [16, §11] and Chung-Lu graphs [9] in certain regimes. These regimes where this bound may be effective are those where there is constant minimum degree, the average degrees is large, and it is hard to get good enough control on the spectral statistics to apply spectral methods to obtain estimates on the resistances or hitting times with the correct leading constant.

We also consider expected hitting times $h(i,j)$ of a random walk. Let $\mathbf{P}_i^G$ be the law of a simple random walk (SRW) X_t on G, that is the random process which at each step moves to a uniformly chosen neighbour of the current vertex, then $h(i,j) := \mathbf{E}^G t[\tau_j | X_0 = i]$, where $\tau_j := \inf\{t : X_t = j\}$. Hitting times are well studied in Markov chain theory [1, 22]. They also feature in randomised algorithms, for example, the run time of the original LOGSPACE algorithm for undirected complexity [23], and are a popular tool in machine learning to analyze the structure of graphs [27]. Tetali's formula [26] relates hitting times to resistances:

$$h(i,j) = |E(G)| \cdot R(i,j) + \sum_{u \in V} \frac{d(u)}{2}[R(j,u) - R(u,i)]. \tag{3}$$

Using Tetali's formula we derive results for hitting times and related qualities via controlling resistances. We shall focus on the following regime for $\mathcal{G}(n,p)$ which we call sparsely connected:

$$\log n + \log\,\log\,\log n \le np \le n^{1/10}. \tag{4}$$

Recall that $\mathcal{G}(n,p)$ has average degree np, however, at the lower end of the range (4) it also has vertices of constant degree w.h.p. so, at the lower end, $\mathcal{G}(n,p)$ is far from being regular.

Let $\mathcal{C} := \mathcal{C}_n$ be the event $\mathcal{G}$ is connected and $\mathbb{E}[\cdot|\mathcal{C}]$ be conditional expectation w.r.t. $\mathcal{G}(n,p)$.

Theorem 1.2. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ satisfy (4). Then for any $i,j \in V(\mathcal{G})$, where $i \neq j$,*

$$\mathbb{E}[R(i,j)|\mathcal{C}] = \frac{2 \pm O(\varepsilon_n)}{np} \quad \textit{and} \quad \mathbb{E}[h(i,j)|\mathcal{C}] = n(1 \pm O(\varepsilon_n)).$$

We obtain concentration for resistances and hitting times from Theorem 1.1.

Theorem 1.3. *For any $c > 0$ let $\mathcal{G} \sim \mathcal{G}(n,p)$ with $np = (c \pm o(1))\log n$. Then for fixed $i \neq j \in V$,*

(*i*) $\mathbb{P}\left(\left|R(i,j) - \frac{2}{np}\right| > \frac{10}{c^2 \log(n)\log\,\log(n)}\right) \le e^{-\Omega\left(\frac{\log n}{(\log\,\log n)^2}\right)}$. *Further, if $np = \omega(\log n)$ and $np \le n^{1/10}$ then*

(*ii*) $\mathbb{P}\left(\sup_{\{i,j\} \subseteq V}\left|R(i,j) - \frac{2}{np}\right| > \frac{7\sqrt{\log n}}{(np)^{3/2}}\right) = o\left(\frac{1}{n}\right)$.

(*iii*) $\mathbb{P}\left(\sup_{\{i,j\} \subseteq V}|h(i,j) - n| > 12n\sqrt{\frac{\log n}{np}}\right) = o\left(\frac{1}{n}\right)$.

Observe Theorem 1.3 (iii) gives concentration of $h(i,j)$ around n for all pairs $i,j \in V$ when $np = \omega(\log n)$. For $np = \Theta(\log n)$ we prove concentration by the second moment method.

Theorem 1.4. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ satisfy (4), $f(n) : \mathbb{N} \to \mathbb{R}_+$. Then for a fixed $i, j \in V$, $i \neq j$,*

$$\mathbb{P}(|h(i,j) - n| > n\sqrt{f(n) \cdot \varepsilon_n}) = O\left(\frac{1}{f(n)}\right).$$

In particular by choosing $f(n) = \log\,\log(np)$ above we have concentration for a fixed pair $i, j \in V$, not all pairs; the following proposition shows that this is best possible.

Proposition 1.5. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$. If $np = \log n + 100 \log\,\log\,\log n$, then w.h.p. there exists $i, j \in V$ such that $R(i,j) \geq 1$ and $h(i,j) > n\,\log(n)/3$. For any $1 < c < \infty$, if $np = c\,\log(n)$ then there is an $a > 0$ and $i, j \in V$ such that w.h.p. $R(i,j) \geq (2+a)/np$ and $h(i,j) > (1+a)n$.*

Theorems 1.2 to 1.4 are valid only for $np \leq n^{1/10}$, however, concentration and expectation for all of the aforementioned random variables has been determined for np above this range. The original contribution of this paper is determining expectation and concentration close to the connectivity threshold $np = \log n$, see the literature review in Section 1.2 for more details.

One consequence of applying Theorem 2.3 to $\mathcal{G}(n,p)$ is that we can also show that there are many ways to select a edge-disjoint paths between the first neighbours of a pair of vertices. In particular for a graph G let $paths_2(i,j,l)$ be the maximum number of paths of length at most l between vertices i and j of G that are vertex disjoint on $V \backslash (B_1(i) \cup B_1(j))$.

Theorem 1.6. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, where for any $c > 0$, $c\,\log n \leq np \leq n^{1/10}$. Let $l := \log n/\log(np) + 9$. Then for $i, j \in V$, where $i \neq j$,*

(i) $\mathbb{P}(paths_2(i,j,l) \neq \min\{d_2(i), d_2(j)\}) \leq 5n^3p^4 + o(e^{-7\min\{np, \log n\}/2})$,
(ii) $\mathbb{P}(|paths_2(i,j,l) - (np)^2| > 3(np)^{3/2}\sqrt{\log np}) = o(1/np)$.

We also prove results for some other related indices which appear in the literature for $\mathcal{G}(n,p)$. For a discussion of how our results extend previous work see Section 1.2.

Let $\pi(v) = d(v)/2|E|$ for $v \in V$ be the stationary distribution of the SRW on G and define,

$$H_j(G) := \sum_{i \in V} \pi(i)h(i,j) \quad \text{for} \quad j \in V, \qquad T(G) := \sum_{j \in V} \pi(j)h(i,j). \tag{5}$$

The index $H_j(G)$ is known as the stationary hitting time to j [24] and $T(G)$ is the random target time or Kemeny's constant [1,22]. Note that $T(G)$ is independent of the vertex i in (5), see [23, Equation 3.3], and the expected running time of Wilson's algorithm [28] on G is $O(T(G))$.

Theorem 1.7. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ satisfy (4). Then for any $i \in V(\mathcal{G})$,*

$$\mathbb{E}[H_i(\mathcal{G})|\mathcal{C}] = n(1 \pm O(\varepsilon_n)) \quad \text{and} \quad \mathbb{E}[T(\mathcal{G})|\mathcal{C}] = n(1 \pm O(\varepsilon_n)).$$

The Kirchoff index $K(G)$ and cover cost $cc_i(G)$ of a finite connected graph G are defined by

$$K(G) := \sum_{\{i,j\} \subseteq V} R(i,j), \quad \text{and} \quad cc_i(G) := \frac{1}{n-1}\sum_{j \in V} h(i,j) \quad \text{for} \quad i \in V. \tag{6}$$

The former is studied in the contexts of mathematical chemistry [12] and sensor networks [5], and the latter was introduced to bound the cover time [17,18]. By linearity of expectation:

Corollary 1.8 (Of Theorem 1.2). *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ satisfy (4). Then for any $i \in V(\mathcal{G})$,*

$$\mathbb{E}[K(\mathcal{G})|\mathcal{C}] = \frac{n}{p}(1 \pm O(\varepsilon_n)) \quad \text{and} \quad \mathbb{E}[cc_i(\mathcal{G})|\mathcal{C}] = n(1 \pm O(\varepsilon_n)).$$

We prove concentration for these random variables on $\mathcal{G}(n,p)$ by the second moment method.

Theorem 1.9. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ satisfy (4) and let $f(n) : \mathbb{N} \to \mathbb{R}_+$. Fix $i \in V$ and let X be any of the random variables $K(\mathcal{G})$, $H_i(\mathcal{G})$, $T(\mathcal{G})$, $cc_i(\mathcal{G})$. Then*

$$\mathbb{P}(|X - \mathbb{E}[X\,|\,\mathcal{C}]| > \mathbb{E}[X\,|\,\mathcal{C}]\sqrt{f(n)\cdot\varepsilon_n}) = O\left(\frac{1}{f(n)}\right).$$

1.1 | Outline of the Paper

Section 2 contains our bounds on the effective resistance. In particular, in Section 2.1 we prove a general bound, Theorem 2.3, which is based on the existence and structure of a desirable subgraph H. In Section 2.2 we describe a specialisation of this bound based on a specific family of subgraphs defined by an exploration process which is well suited to $\mathcal{G}(n,p)$. In Section 3 we prove preliminary results regarding the subgraph of $\mathcal{G}(n,p)$ described in Section 2.2, these results are needed to apply our bounds on effective resistance to $\mathcal{G}(n,p)$. In Section 4 we apply the results of Sections 2 and 3 to prove Theorems 1.1 and 1.3, which determine resistances in $\mathcal{G}(n,p)$, and Theorem 1.6, which concerns paths between second neighbours in $\mathcal{G}(n,p)$. Finally in Section 5 we combine results from the previous three sections to prove Theorems 1.2, 1.4, 1.7, and 1.9 which are results on the expectation and concentration for hitting times of random walks and related indices on $\mathcal{G}(n,p)$. In the remainder of this section we shall discuss some related work and how our work extends known results and also cover some preliminary material.

1.2 | Related work

In [20] Jonasson studies the cover time, that is the expected time to visit all vertices from the worst starting vertex, for $\mathcal{G}(n,p)$. He bounds the cover time by showing effective resistances and hitting times on $\mathcal{G}(n,p)$ concentrate in the regimes, where $\omega(\log n) = np \leq n^{1/3}$. Jonasson does not use spectral methods and instead bounds the effective resistance by finding a suitable flow. This is the approach we have also taken, using a refined analysis we extend Jonasson's results for hitting times to the range (4) and for effective resistance to the case $np = \Omega(\log n)$. It is worth noting that the cover time has since been determined for all connected $\mathcal{G}(n,p)$ by Cooper and Frieze [11] using the first visit Lemma and mixing time estimates.

Let $L = D - A$ be the graph Laplacian, where A is the adjacency matrix and D is the diagonal matrix with $D_{i,j} = d(i)$ if $i = j$ and $D_{i,j} = 0$ otherwise [12, 23]. Many previous results rely on exploiting connections between resistances or hitting times and spectral statistics of L or other representations of the graph. In this paper we do not employ spectral methods; the results we achieve hold for $\mathcal{G}(n,p)$ close to the connectivity threshold where the minimum degree is 1 w.h.p. and it is hard to obtain good enough estimates on the relevant spectral statistics.

Boumal and Cheng [5] exploit an expression for the Kirchoff index $K(G)$ in terms of the trace of $L^{\dagger}(G)$, the Moore-Penrose pseudoinverse of $L(G)$ [12], to obtain expectation and concentration for $K(\mathcal{G})$ on $\mathcal{G}(n,p)$ with $np = \omega((\log n)^6)$. We will now outline a related expression for $K(G)$ and explain how this can also be used with spectral statistics to control $K(G)$. Let λ_i be the eigenvalues of $L(G)$, where G is a finite connected graph. Then by the matrix tree theorem [18]:

$$K(G) = \sum_{\lambda_i \neq 0} \frac{1}{\lambda_i}. \tag{7}$$

A theorem of Coja-Oghlan [10, Theorem 1.3], states that if $\mathcal{G} \sim \mathcal{G}(n,p)$ with $np \geq C_0 \log n$ for sufficiently large C_0 the nonzero eigenvalues of $L(\mathcal{G})$ concentrate around the mean. Combining these estimates with (7) yields concentration for $K(\mathcal{G})$ and with extra work the leading order term of $\mathbb{E}[K(\mathcal{G})|\mathcal{C}]$ can be determined when $np \geq C_0 \log n$. It is of note however that Boumal and Cheng obtain second order terms for $\mathbb{E}[K(\mathcal{G})|\mathcal{C}]$, which is not possible with the latter method. Theorems 1.2 and 1.9 give expectation and concentration for $K(\mathcal{G})$ in the range (4).

Löwe and Torres [24] obtain concentration results for $T(\mathcal{G}), H_i(\mathcal{G})$ and also the commute time $\kappa(i,j) = h(i,j) + h(j,i)$ on $\mathcal{G}(n,p)$. Again, the result comes from using expressions for these quantities in terms of the eigenvectors and eigenvalues of the transition matrix of the simple random walk, these expressions can be found in [23]. Löwe and Torres then apply results from Erdős et al [14] to bound from above the reciprocal of the spectral gap. Löwe and Torres require $np = \omega((\log n)^{C_0})$ for some $C_0 > 0$ sufficiently large as this is needed to apply the results in [14]. Theorems 1.2, 1.7, 1.9, and 1.4 extend these results to the range (4).

Von Luxburg et al [27, Theorem 5] prove bounds on the difference of $h(i,j)/2|E|$ from $1/d(i)$ for nonbipartite graphs by the reciprocal of the spectral gap and the minimum degree of G. They then apply these to various geometric random graphs. These bounds give the same result as Theorem 1.1(iii) when applied to $\mathcal{G}(n,p)$ with $np = \omega(\log n)$, however, if $np = O(\log n)$ they will only give the hitting times up-to a constant. Theorem 1.4 provides concentration results for $h(i,j)$ recovering the leading constant in the extended range (4).

On a different note, Bollobás and Thomason [4, Theorem 7.4] showed the threshold for having minimum degree $k(n)$ coincides with the threshold for having at least $k(n)$ vertex-disjoint paths between any two points. Theorem 1.6 can be thought of as a "local first neighbourhood relaxation" of this statement for two vertices as it roughly states that if you want to separate two vertices x and y and your not allowed to use any vertices from either x or y's first neighbourhoods then w.h.p. the next best option take the smaller of x or y's second neighbourhoods as a separator. Broder et al [6] show that there are edge disjoint paths between any two sets of vertices in $\mathcal{G}(n,p)$, provided that the sets are not too large and provide a polynomial time algorithm to find them. The restrictions on the sets are very modest, however, their results do not give bounds on the length of the paths found or exact bounds on their number.

1.3 | Further preliminaries

We use $X \sim \mathcal{L}$ to denote the random variable X having law $\mathcal{L}$. For random variables A, B, we say that B dominates A if $\mathbb{P}[A > x] \leq \mathbb{P}[B > x]$ for every x and we use the notation $B \succcurlyeq_1 A$, or $A \preccurlyeq_1 B$ in this case. If $A \preccurlyeq_1 B$ and $A, B \geq 0$ then $\mathbb{E}[A^{\alpha}] \leq \mathbb{E}[B^{\alpha}]$ for any $\alpha \geq 1$. Let $\mathrm{Bin}(n,p)$ denote the binomial distribution over n trials each of probability p. Some additional probabilistic notions and lemmas may be found in Appendix A.

Throughout we will be working on a finite simple connected graph $G = (V, E)$ with $|V| = n$ and $|E| =: m$. Let $d(i,j)$ be the graph distance between $i, j \in V$ and define the following:

$$\Gamma_{G,k}(i) := \{j \in V : d(i,j) = k\}, \quad d_{G,k}(i) := |\Gamma_k(i)|, \quad B_{G,k}(i) := \bigcup_{h=0}^{k} \Gamma_h(i), \tag{8}$$

which are the kth neighbourhood of i, size of kth neighbourhood and the ball of radius k centred at i, respectively. We drop the G from the subscripts in (8) when the graph is clear, and the subscript 1 when referring to first neighbourhoods, that is, $\Gamma(x) := \Gamma_1(x)$ and $d(x) := |\Gamma(x)|$.

The hitting times $h(i,j)$ can be far from symmetric, see the example of the lollipop graph [23]. The commute time $\kappa(i,j)$ is the expected number of steps for a random walk from i to reach j and return back to i. The commute time $\kappa(i,j)$ is symmetric and related to hitting times and effective resistances by the commute time formula [7]

$$\kappa(i,j) := h(i,j) + h(j,i) = 2m \cdot R(i,j). \tag{9}$$

1.3.1 | Erdős-Rényi Graphs

The Erdős-Rényi or Binomial random graph model $\mathcal{G}(n,p)$ is a probability distribution over simple n vertex graphs. Any given n vertex graph $G = (V, E)$ is sampled with probability

$$\mathbb{P}(\mathcal{G} = G) = p^{|E(G)|}(1-p)^{\binom{n}{2} - |E(G)|}.$$

This $\mathbb{P}$ is the product measure over edges of the complete graph K_n, where each edge occurs as an i.i.d. Bernoulli random variable with probability $0 < p := p(n) < 1$. Throughout $\mathbb{E}$ will denote expectation with respect to $\mathbb{P}$. Another feature of Erdős-Rényi graphs worth mentioning is that for each $u \in V$ the degree of u is binomially distributed $d(u) \sim Bin(n-1, p)$ and the degrees are not independent. This model has received near constant attention in the literature since the original $\mathcal{G}(n,m)$ model was studied by Erdős and Rényi [15]. For more information consult one of the many books on random graphs [4, 16, 19].

Observe that the effective resistance becomes a random variable when the graph is drawn from $\mathcal{G}(n,p)$. Since the effective resistance between two disconnected vertices is infinite we shall need to condition on the event $\mathcal{C} := \mathcal{C}_n$ that $\mathcal{G}$ is connected. Let $\mathbb{P}_{\mathcal{C}}(\cdot) := \mathbb{P}(\cdot|\mathcal{C})$ and $\mathbb{E}_{\mathcal{C}} := \mathbb{E}[\cdot|\mathcal{C}]$ be the expectation with respect to $\mathbb{P}_{\mathcal{C}}$. The following theorem gives a bound on being disconnected above the $np = \log n$ connectivity threshold.

Theorem 1.10 (Bollobás [3, Theorem 9, §VII]). *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, $np = \log n + \omega(n)$, where $\omega(n) \to \infty$. Then*

$$\mathbb{P}(\mathcal{C}^c) \leq 4 \cdot e^{-\omega(n)}. \tag{10}$$

1.3.2 | Basics of electrical network theory

There is a rich connection between random walks on graphs and electrical networks, consult either of the books [13, 25] for a thorough introduction to the subject. Here we only intend to make the definition of $R(i,j)$ given in the introduction rigorous and cover the essentials.

An electrical network, $N := (G, C)$, is a graph G and an assignment of conductances $C : E(G) \to \mathbb{R}^{+}$ to the edges of G. Our graph G is undirected and we define $\vec{E}(G) := \{\overrightarrow{xy} : xy \in E(G)\}$, this is the set of all possible oriented edges for which there is an edge in G. For some $i, j \in V(G)$, a flow from i to j is a function $\theta : \vec{E}(G) \to \mathbb{R}$ satisfying $\theta(\overrightarrow{xy}) = -\theta(\overrightarrow{yx})$ for every $xy \in E(G)$ as well as Kirchoff's node law for every vertex apart from i and j, that is

$$\sum_{u \in \Gamma_1(v)} \theta(\overrightarrow{uv}) = 0 \quad \text{for each } v \in V,\ v \neq i, j.$$

A flow from i and j is called a unit flow if in addition to the above it has strength 1, that is

$$\sum_{u \in \Gamma_1(i)} \theta(\overrightarrow{iu}) = 1, \qquad \sum_{u \in \Gamma_1(j)} \theta(\overrightarrow{uj}) = 1.$$

For the network $N = (G, C)$ we can then define the effective resistance $R_C(i, j)$ between two vertices $i, j \in V(G)$. First for a flow θ on N let

$$\mathcal{E}(\theta) = \sum_{e \in \vec{E}} \frac{\theta(e)^2}{2C(e)},$$

be the energy dissipated by θ. Then for $i, j \in V(G)$, $R_C(i, j)$ can be defined as

$$R_C(i, j) := \inf\{\mathcal{E}(\theta) : \theta \text{ is a unit flow from } i \textit{ to } j\}. \tag{11}$$

We refer to the flow minimising (11) as the current, thus the effective resistance is the energy dissipated by the current of strength 1 from i to j in $N = (G, C)$. This current exists and is unique since we are working on a finite graph. Equivalently $R(i, j)$ is the reciprocal of the amount of flow when a unit potential difference is fixed between i and j.

The conductances C define a reversible Markov chain [25]. In this paper we fix $C(e) = 1$ for all $e \in E(G)$ as this corresponds to a simple random walk, in this case we write $R(i, j)$ instead of $R_C(i, j)$. This $R(i, j)$ is the effective resistance in Equations (6), (3), and (9).

One very useful tool is Rayleigh's monotonicity law [25, §2.4]: If $C, C' : E(G) \to \mathcal{R}^{+}$ are conductances on the edge set $E(G)$ of a connected graph G and $C(e) \leq C'(e)$ for all $e \in E(G)$ then for all pairs $\{i, j\} \subset V(G)$, we have $R_{C'}(i, j) \leq R_C(i, j)$.

2 | BOUNDS ON EFFECTIVE RESISTANCE

The aim of this section is to obtain bounds on $R(u, v)$ for which the main contribution, when applied to graphs with good expansion, comes from the first neighbourhoods of u and v.

2.1 | General bound

Recall that $d(x)$ denotes the size of the first neighbourhood of vertex $x \in V(G)$. Jonasson gives the following lower bound on effective resistance.

Lemma 2.1 (Jonasson [20, Lemma 1.4]). *For any graph $G = (V, E)$ and $x, y \in V$, $x \neq y$*

$$R(x, y) \geq \frac{1}{d(x) + 1} + \frac{1}{d(y) + 1}.$$

We seek an upper bound where the dominant term looks roughly like the one in Lemma 2.1. To achieve this we shall define a subgraph $H(x, y, r, k)$ which will allow us to route flow through the graph efficiently. Recall the definition (8) of $\Gamma_{H,k}(x)$, $d_{H,k}(x)$, and $B_{H,k}(x)$.

Definition 2.2. For a graph G and $x, y \in V(G)$, $x \neq y$ define $H := H(x, y, r, k) \subseteq G$ as follows. First, $B_{H,r}(x)$ (resp. $B_{H,r}(y)$) is a tree with a nonempty set of leaves all at distance r from x (resp. y). Second, for each vertex $w \in \Gamma_{H,r}(x) \cup \Gamma_{H,r}(y)$ there exists a set H_w such that

(i) The vertex w is connected to every vertex in H_w by paths of lengths at most k.
(ii) For every distinct $w, z \in \Gamma_{H,r}(x) \cup \Gamma_{H,r}(y)$, the sets H_w, H_z are disjoint and the paths connecting w to H_w and z to H_z are edge disjoint.
(iii) For every $(w, z) \in \Gamma_{H,r}(x) \times \Gamma_{H,r}(y)$ there is an edge from H_w to H_z.

To better understand why defining $H(x, y, r, k)$ as above will help us control effective resistance $R(x, y)$ in graphs with good expansion we first recall definition (11) which states that $R(x, y)$ is determined by the sum over all edges of the square of the current through each edge. Heuristically one wishes to keep the max flow through an edge as small as possible, since this max is squared. If a graph has good expansion properties, that is every set has lots of edges leaving the set, then the smallest edge cuts separating x from y will be "close" to x and y. Our approach is thus to find a flow from x to y which balances the flow as evenly as possible over edges close to x and y where the cuts are smaller and thus the amount of flow is greater. We divide the flow evenly among all descendants of x (from the perspective of rooting $B_{H,r}(x)$ at x) and do the same for the flows in the reverse direction in $B_{H,r}(y)$. Then once we are clear of the r-neighbourhoods we use the abundance of paths to route the flow from $B_{H,r}(x)$ to $B_{H,r}(y)$. The challenge is to then make sure all the paths meet up and that we have a valid flow, the structure of the desired subgraph H outlined in Definition 2.2 will allow us to do this.

We now bound $R(x, y)$ in terms of the neighbourhoods of vertices in $H(x, y, r, k)$. This bound is effective when a subgraph H can be found who's vertices close to x and y have degrees similar to what they were in the original graph. Let $\mathcal{P}_i(x) := \{x_0x_1 \cdots x_i : x_0 = x,\ x_{k-1}x_k \in E(H)\}$ be the set of all paths of length i from x in H, and $\mathbf{1}_{j=x} = 1$ if $j = x$ and 0 otherwise.

Theorem 2.3. *For a graph G, $x, y \in V$ disjoint and $k, r \geq 1$, if $H(x, y, r, k) \subseteq G$ then*

$$R(x, y) \leq \frac{1}{d_H(x)} + \frac{1}{d_H(y)} + \sum_{i=1}^{r} \sum_{pp_1\cdots p_i} \frac{1}{d_H(p)^2} \prod_{j=1}^{i-1} \frac{1 + \mathbf{1}_{j=r-1} \cdot (k+1)}{(d_H(p_j) - 1)^2},$$

where summation $pp_1 \cdots p_i \in \mathcal{P}_i(x) \cup \mathcal{P}_i(y)$ is over all paths of length i from x or y in H.

Proof. We will now describe a unit flow θ from x to y through the network $N = (H, C)$, where $C(e) = 1$ for all $e \in E(H)$. This flow will be used to bound $R(x, y)$ from above by (11).

To begin we assign a flow of $\theta(xx_1) = 1/d_H(x)$, where $x_1 \in \Gamma_H(x)$ is a neighbour of x. Likewise let $\theta(yy_1) = -1/d_H(y)$, where $y_1 \in \Gamma_H(y)$. Then, for each edge $x_{i-1}x_i$, where $x_i \in \Gamma_{H,i}(x)$ and $1 \le i \le r$ we send the amount of flow entering x_i divided by the number of edges to $x_i \in \Gamma_{H,i}(x)$. So inductively if the unique path from $x = x_0$ to some x_i is $x_0, x_1, \ldots, x_i$, then the flow through the (directed) edge $x_{i-1}x_i$ is

$$\theta(x_{i-1}x_i) = \frac{1}{d_H(x)(h(x_1)-1)\cdots(d_H(x_{i-1})-1)} = \frac{1}{d_H(x)}\prod_{j=1}^{i-1}\frac{1}{d_H(x_j)-1},$$

where we follow the convention that empty products are equal to 1.

We do the same with edges in the r neighbourhood of y but the flow is reversed. The total contribution to $\mathcal{E}(\theta)$ from the ball $B_{H,r}(x)$ of radius r around x is then given by

$$\sum_{i=1}^{r}\sum_{xx_1\cdots x_i\in\mathcal{P}_i(x)}\frac{1}{d_H(x)^2}\prod_{j=1}^{i-1}\frac{1}{(d_H(x_j)-1)^2} \tag{12}$$

and likewise for the contribution to $\mathcal{E}(\theta)$ from the edges in $B_{H,r}(y)$.

We now describe the flow across an edge from H_w to H_z, where $(w,z) \in \Gamma_{H,r}(x) \times \Gamma_{H,r}(y)$. Indeed for each such edge $e_{w,z}$ we assign a flow

$$\theta(e_{w,z}) = \theta(x_{r-1}w)\cdot\theta(zy_{r-1}) = \frac{1}{d_H(x)\cdots(d_H(x_{r-1})-1)}\cdot\frac{1}{d_H(y)\cdots(d_H(y_{r-1})-1)},$$

where $xx_1\cdots x_{r-1}w \in \mathcal{P}_r(x)$ (resp. $yy_1\cdots y_{r-1}z \in \mathcal{P}_r(y)$) is the unique path of length r from x (resp. y) to w (resp. z) in H. The reason for assigning this flow is that if we sum over all the vertices in $\Gamma_{H,r}(y)$ we obtain

$$\sum_{z\in\Gamma_{H,r}(y)}\theta(e_{w,z}) = \theta(x_{r-1}w)\sum_{\mathcal{P}_r(y)}\frac{1}{d_H(y)\cdots(h(y_{r-1})-1)} = \theta(x_{r-1}w), \tag{13}$$

which is precisely the flow leaving through $w \in \Gamma_{H,r}(x)$. Thus the contribution to $\mathcal{E}(\theta)$ by the flow through these edges is precisely

$$\sum_{xx_1\cdots x_r\in\mathcal{P}_r(x)}\frac{1}{d_H(x)^2}\prod_{j=1}^{r-1}\frac{1}{(d_H(x_j)-1)^2} = \sum_{yy_1\cdots y_r\in\mathcal{P}_r(y)}\frac{1}{d_H(y)^2}\prod_{j=1}^{r-1}\frac{1}{(d_H(y_k)-1)^2} \tag{14}$$

We are not concerned with how flow is rooted from w to the relevant vertices of H_w but note that since there is a path from w to each vertex in H_w with an edge to some H_z, where $z \in \Gamma_{H,r}(y)$ constructing a flow is possible. We now bound the contribution from these paths.

Claim. The contribution to the $\mathcal{E}(\theta)$ by the flow through the paths from $w \in \Gamma_{H,r}(x)$ to H_w is at most $k\cdot\theta(x_{r-1}w)^2$, where $xx_1\cdots x_{r-1}w \in \mathcal{P}_r(x)$. The analogous bound holds for $z \in \Gamma_{H,r}(y)$.

Proof of claim. We can assume that all paths to vertices in H_w have length k otherwise we can subdivide edges on these paths, only increasing total resistance by Rayleigh's monotonicity law. Consider the set S_i of edges in the union of all paths to H_w with furthest endpoint from w at distance $1 \le \ell \le k$ from w, this edge set separates w from H_w.

Recalling property (ii) of Definition 2.2, the only flow through these edges is that from w to H_w. Thus the combined flow through S_i is $\theta(x_{r-1}w)$ since this is the amount of flow entering at w and leaving H_w, as shown by (13). Thus since the contribution to $\mathcal{E}(\theta)$ by the edges of S_i is the sum of the squares of the flows through each edge of S_i we see that this cannot exceed $\theta(x_{r-1}w)^2$ by convexity. The result follows by summing the contributions from the k such edge sets S_i. □

Thus the contribution to $\mathcal{E}(\theta)$ from all edges in these paths is at most

$$\sum_{xx_1\cdots x_r\in\mathcal{P}_r(x)}\frac{k}{d_H(x)^2}\prod_{j=1}^{r-1}\frac{1}{(d_H(x_j)-1)^2}+\sum_{yy_1\cdots y_r\in\mathcal{P}_r(y)}\frac{k}{d_H(y)^2}\prod_{j=1}^{r-1}\frac{1}{(d_H(y_j)-1)^2}. \tag{15}$$

The result follows by summing the contributions (12), (14), and (15) to $\mathcal{E}(\theta)$. □

2.2 | Application to $\mathcal{G}(n, p)$

To apply Theorem 2.3 to $\mathcal{G}(n, p)$ one must describe a suitable $H(x, y, r, k)$, this is achieved using the modified breadth-first search (MBFS) algorithm. The inputs to the MBFS algorithm are a graph G and a subset $I_0 = \{u, v\} \subseteq V(G)$, the outputs are sets $I_i, S_i \subseteq V(G)$ and $E_i \subseteq E(G)$ indexed by the iteration of the algorithm. The algorithm is similar to one used in [2, §11.5] to explore the giant component of an Erdős-Rényi graph. However, the MBFS algorithm differs from other variations on breadth-first search algorithms used in the literature as it starts from two distinct vertices. More importantly it also differs by removing clashes, where a clash is a vertex with more than one parent in the previous generation as exposed by a breadth-first search from two root vertices. In what follows all graphs are on a common labelled vertex set $V := [n]$.

Modified breadth-first search algorithm, MBFS(G, I_0): To begin set $S_0 := V\backslash I_0$, and $I_i = E_i = \emptyset$ for all $i \geq 1$. Then generate the sets S_i and update the sets I_i and E_i for $i \geq 1$ iteratively by the following procedure:

*Step*1: Set $S_i = S_{i-1}$. For each $w \in S_i$ check all pairs $\{w, w'\}$, where $w' \in I_{i-1}$ and,
- if there exists $w' \in I_{i-1}$ such that $ww' \in E(G)$ then remove w from S_i,
- if there is a unique $w' \in I_{i-1}$ such that $ww' \in E(G)$ then add w to I_i and add ww' to E_i.

Step 2: If $S_i \neq \emptyset$ and $I_i \neq \emptyset$ then advance i to $i + 1$ and return to step 1. Otherwise end.

The set I_i contains the “active” vertices in the ith iteration and S_i is the set of vertices that have not been used in the first i iterations and E_i is the set $\{xy \in E(G) : x \in I_{i-1}, y \in I_i\}$ of edges “accepted” by the algorithm. Notice that $S_0 \supseteq S_1 \supseteq S_2$... and the sets $\{I_i\}_{i\geq 0}$ are all disjoint. A vertex in S_i will not be included in either I_{i+1} or S_{i+1} if it has two or more neighbourhoods in I_i, in this instance it is just ignored by the algorithm. If instead those vertices in S_i with edges to more than one vertex in I_i were added to I_{i+1} then this procedure would describe a standard breadth-first search starting from two root vertices. Notice also that in step 1 the order in which we consider the vertices of S_i and then the edges between S_i and I_i is unimportant.

For each pair of vertices $I_0 \subseteq V$ the MBFS algorithm provides a filtration

$$\mathfrak{F}_i := \mathfrak{F}_i(I_0), \tag{16}$$

where $\mathfrak{F}_0 \subseteq \mathfrak{F}_1 \subseteq \cdots$, on the set of labelled graphs on V. Roughly speaking $\mathfrak{F}_i(I_0)$ only sees graphs that are the distinguishable by MBFS run up to step $i \geq 0$ from initial set I_0. To make this precise we must first describe an equivalence relation on graphs. Let $u, v \in V$ and G, F be graphs on V. We say $G \cong_k^{\{u,v\}} F$ if the same k-sequence of sets $\{S_i, I_i, E_i\}_{1 \leq i \leq k}$ is output when MBFS$(G, \{u, v\})$ and MBFS$(F, \{u, v\})$ are run for k iterations. Let $I_0 = \{u, v\} \subseteq V$ and define $\mathfrak{F}_i(I_0)$ to be the σ-algebra where the atoms are the equivalence classes of $\cong_i^{\{u,v\}}$.

Let $x \in I_k$, where I_k is produced by running MBFS(G, I_0) for some given I_0. We shall now define $\Gamma_i^*(x)$, the MBFS neighbourhood of x, let $\Gamma_0^*(x) := x$ and for $i \geq 1$

$$\Gamma_i^*(x) := \left\{ y \in I_{k+i} : \begin{array}{c} \text{there exists } x = x_0, x_1, \ldots, x_i = y, \text{ where} \\ x_{j-1}x_j \in E_{k+j} \text{ for all } j = 1, \ldots, i \end{array} \right\}, \tag{17}$$

and let $d_i^*(x) = |\Gamma_i^*(x)|$. Equivalently, we can also define $\Gamma_i^*(x)$ for $i \geq 1$ inductively

$$\Gamma_i^*(x) := \{z \in I_{k+i} : \text{ there exists } y \in \Gamma_{i-1}^*(x) \quad \text{and} \quad yz \in E_i\}.$$

To try and further clarify (17) we define the following sets $S_k(x)$ which are the vertices in S_k that will not cause any clashes when the Γ^*-neighbourhood of x is explored,

$$S_k(x) := S_k \Bigg\backslash \left(\bigcup_{z \in I_k,\ z \neq x} \Gamma_1(z) \right). \tag{18}$$

We can then also define the neighbourhood $\Gamma_i^*(x)$ inductively as follows:

$$\Gamma_i^*(x) = \bigcup_{y \in \Gamma_{i-1}^*(x)} \Gamma_1(y) \cap S_{k+i}(y). \tag{19}$$

We define the pruned neighbourhood $\Phi(x)$ of $x \in I_1$ by

$$\Phi(x) := \Gamma_1^*(x) \backslash \{y : \ d^*(y) \leq D\}, \quad \text{and let} \ \ \varphi(x) := |\Phi(x)|, \tag{20}$$

where

$$D := D(n, p) = \begin{cases} \max\left\{ \left\lceil \dfrac{50}{c} \right\rceil, 50 \right\}, & \text{if } np = (c \pm o(1))\log(n), \text{ where } c > 0, \\ 0, & \text{if } np = \omega(\log(n)). \end{cases} \tag{21}$$

This is the MBFS neighbourhood of x with all the neighbours who have less than D "MBFS-children" removed. The choice of D is related to concentration for binomial random variables and our choice for this will be apparent during the proof of Lemma 3.3. One may think of D as an atypically small value for the degree of a vertex.

Define the pruned neighbourhoods $\Psi_1(w)$ of $w \in I_0$ by

$$\Psi_1(w) := \Gamma_1^*(w) \backslash \{y : \Phi(y) = \emptyset\}, \quad \text{and let} \quad \psi(w) := |\Psi_1(w)|. \tag{22}$$

Set $\Psi_0(u) = \{u\}$ and let the pruned second neighbourhood $\Psi_2(w)$ of $w \in I_0$ be given by

$$\Psi_2(w) := \bigcup_{x\in\Psi_1(w)} \Phi(x) = \bigcup_{x\in\Gamma_1^*(w)} \Phi(x), \quad \text{also let} \quad \psi_2(w) := |\Psi_2(w)|. \tag{23}$$

For MBFS($G, \{u, v\}$) define Ψ_i, the pruned version of I_i for $i = 0, 1, 2$, by

$$\Psi_i := \Psi_i(u) \cup \Psi_i(v), \quad i = 0, 1, 2.$$

We prune the first neighbourhoods of vertices $x \in I_1$ to obtain $\Phi(x)$ so that later on when we consider the trees induced by the union up to i of the Γ^*-neighbourhoods of $y \in \Phi(x)$ we can get good control over the growth rate of the trees. We prune the first neighbourhoods of vertices $w \in I_0$ as above so that we can send flow from our source vertex w to its pruned neighbourhood $\Psi_1(w)$ without having to worry about it getting stuck in any "dead ends."

Recall (16), the definition of the filtration $\mathfrak{F}_k(I_0)$. Observe that if $x \in I_k$ then $\Gamma_1^*(x)$ is $\mathfrak{F}_{k+1}$ measurable. It is worth noting, however, that if $y \in I_1$ then $\Phi(y)$ is $\mathfrak{F}_3$ measurable and not necessary $\mathfrak{F}_2$ measurable since $\Phi(y)$ is determined by vertices at distances 2 and 3 from I_0. A consequence of this is that for $w \in I_0$, $\Psi_1(w), \Psi_2(w)$ are both $\mathfrak{F}_3$ measurable as they are both determined by the Φ-neighbourhoods of vertices in $\Gamma_1^*(w)$.

Definition 2.4 (The set $A_{u,v}^{n,k}$). For integers $n, k \geq 0$ let $A_{u,v}^{n,k}$ be the set of n-vertex graphs on V, where $u, v \in V$, such that for every pair $(x, y) \in \Psi_2(u) \times \Psi_2(v)$ the neighbourhoods $\Gamma_k^*(x)$ and $\Gamma_k^*(y)$ are nonempty and there is at least one edge $ij \in E(G)$, where $i \in \Gamma_k^*(x)$, $j \in \Gamma_k^*(y)$.

We relate the structure of $G \in A_{u,v}^{n,k}$ to Theorem 2.3 to give a bound on $R(u, v)$.

Corollary 2.5 (Of Theorem 2.5). *Run* MBFS($G, \{i, j\}$) *and suppose* $G \in A_{i,j}^{n,k}$. *Then*

$$\begin{aligned} R(i,j) &\leq \frac{1}{\psi(i)} + \frac{1}{\psi(j)} + \sum_{a\in\Psi_1(i)} \frac{k+2}{\psi(i)^2 \cdot \varphi(a)} + \sum_{b\in\Psi_1(j)} \frac{k+2}{\psi(j)^2 \cdot \varphi(b)} \\ &\leq \frac{1}{\psi(i)}\left(1 + \sup_{x\in\Psi_1(i)} \frac{k+2}{\varphi(x)}\right) + \frac{1}{\psi(j)}\left(1 + \sup_{y\in\Psi_1(j)} \frac{k+2}{\varphi(y)}\right). \end{aligned}$$

Proof. If we can find a some suitable subgraph $H(i, j, 2, k)$, from Definition 2.2, encoded by the property $A_{i,j}^{n,k}$ then the result follows by Theorem 2.3. The only thing that can go wrong with $G \in A_{i,j}^{n,k}$ according to Definition 2.4 is that if one of the neighbourhoods $\Psi_1(i)$, $\Psi_1(j)$, $\Psi_2(i)$, or $\Psi_2(j)$ are empty. In this case one of the terms $\psi(i)$,$\psi(i)$, $\varphi(a)$, or $\varphi(b)$ on the RHS of the inequality will be 0, we define 1/0 to be infinity and so the inequality holds vacuously. □

We encode Definition 2.4 as the following event for $\mathcal{G} \sim \mathcal{G}(n, p)$,

$$\mathcal{A}_{u,v} := \left\{\text{exists } k \leq \log n/(2 \log np) + 2 \text{ such that } \mathcal{G} \in A_{u,v}^{n,k}\right\}, \tag{24}$$

and say that $\mathcal{G}$ satisfies the strong path property if this holds, for an illustration of this property consult (Figure 1).

Notice that in Definition 2.4 either $\Psi_1(u)$ or $\Psi_1(v)$ may be empty, thus we also define the following sets $B_w^{u,v}$ for $w \in \{u, v\}$ using the output of MBFS$(G, \{u, v\})$:

$$B_w^{u,v} := \{G : \Psi_1(w) \neq \emptyset\}, \quad \text{and let} \quad B_{u,v} := B_u^{u,v} \cap B_v^{u,v}.$$

Similarly to how we defined $\mathcal{A}_{u,v}$ define the events

$$\mathcal{B}_w^{u,v} := \{\mathcal{G} \in B_w^{u,v}\}, \quad \mathcal{B}_{u,v} = \mathcal{B}_u^{u,v} \cap \mathcal{B}_v^{u,v}. \tag{25}$$

3 | THE STRONG PATH PROPERTY FOR $\mathcal{G}(n, p)$

In this section, we prove some results needed to successfully apply Corollary 2.5 to $\mathcal{G}(n, p)$ in the sparsely connected range (4), in particular we show that $\mathcal{A}_{u,v}$ holds w.h.p. To apply the bound on effective resistance in terms of the reciprocals of ψ and φ we couple them to d and d^*. Lemmas 3.1, 4.2, and 3.4 will help us achieve this.

Lemma 3.1. *Let $\mathcal{G} \sim \mathcal{G}(n, p)$, $I_0 := \{u, v\} \subset V$, and $i, k \geq 0$. Run* MBFS$(\mathcal{G}, I_0)$.

(i) *Then $|S_1| \sim Bin(n-2, (1-p)^2)$ and $|I_1| \sim Bin(n-2, 2p(1-p))$.*

(ii) *Conditioning on $\{x \in I_k\}$ and $|S_k(x)|$, then*

$$d^*(x) \sim Bin(|S_k(x)|, p).$$

(iii) *Conditioning on $\{x \in I_k\}$, $|S_{k+i}|$, $|I_{k+i}|$, and $d_i^*(x)$, then*

$$d_{i+1}^*(x) \sim Bin(|S_{k+i}|, d_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}).$$

Proof. *Item* (i): A vertex in S_0 is in S_1 if it is not connected to either vertex in I_0. This happens independently w.p. $(1-p)^2$ for each of the $n-2$ vertices in S_0 thus $|S_1| \sim Bin(n-2, (1-p)^2)$.

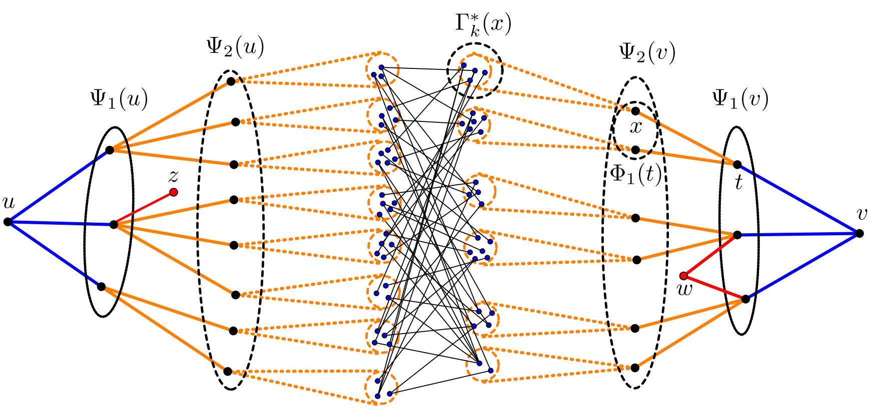

FIGURE 1 Illustration of $G \in A_{u,v}^{n,k}$, see Definition 2.4. Note: vertex z is not in $\Psi_2(u)$ as it is connected to less than D vertices in I_3 and w is not in I_2 as it has more than one parent in I_1 [Color figure can be viewed at wileyonlinelibrary.com]

A vertex in S_0 is in I_1 if it is connected to exactly one vertex in I_0. This happens independently with probability $2p(1-p)$ for each of the $n-2$ vertices in S_0 thus $|I_1| \sim Bin(n-2, 2p(1-p))$.

Item (ii): Recall the definitions of $\Gamma_1^*(x)$ and $S_k(x)$ for $x \in I_k$, given by (17) and (18), respectively. Observe the following relation:

$$\Gamma_1^*(x) = (\Gamma_1(x) \cap S_k) \Big\backslash \bigcup_{y \in I_k,\ y \neq x} \Gamma_1(y) = \Gamma_1(x) \cap S_k(x).$$

Since we completely remove the vertices if they clash, and the edges of $\mathcal{G}$ are independent, the order MBFS explores the neighbourhoods of each $y \in I_k$ is unimportant. Assume that we have explored the neighbourhood of every $y \in I_k$ with $y \neq x$. We then know which vertices in the neutral set S_k will not clash if included in $\Gamma_1(x)$ and these are the vertices in $S_k(x)$. Since edges occur independently with probability p, conditioning on $|S_k(x)|$ yields $d^*(x) \sim Bin(|S_k(x)|, p)$.

Item (iii): For a vertex $v \in S_{k+i}$ we have $v \in \Gamma_{i+1}^*(x)$ when there is exactly one edge $yv \in E(\mathcal{G})$, where $y \in \Gamma_i^*(x)$ and there is no edge of the form $y'v \in E$, where $y' \in I_{k+i}$ and $y' \neq y$. Conditioning on the sizes of I_{k+i} and $\Gamma_i^*(x)$ we see that each $v \in S_{k+i}$ is a member of $\Gamma_{i+1}^*(x)$ with probability $d_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}$. These events are independent, thus

$$d_{i+1}^*(x) \sim Bin(|S_{k+i}|, d_i^*(x) \cdot p \cdot (1-p)^{|I_{k+i}|-1}),$$

conditional on $|S_{k+i}|$, $|I_{k+i}|$, and $d_i^*(x)$. □

Let $x \in I_k$. Choosing $i = 0$ in Lemma 3.1 (iii) gives $d^*(x) \sim Bin(|S_k|, p(1-p)^{|I_k|-1})$ conditional on $|S_k|$ and $|I_k|$ whereas Lemma 3.1 (ii) gives $d^*(x) \sim Bin(|S_k(x)|, p)$ conditional on $|S_k(x)|$. To relate (ii) to (iii) observe that conditional on $|S_k|$ and $|I_k|$, $|S_k(x)| \sim Bin(|S_k|, (1-p)^{|I_k|-1})$. Item (iii) then follows as if $X \sim B(n, p)$ and, conditional on X, $Y \sim B(X, q)$, then Y is a simple binomial variable with distribution $Y \sim B(n, pq)$.

The next two lemmas provide tail estimates on the sizes of Γ_i and Γ_i^*. We prove them by induction where the inductive step comes from applying Chernoff bounds to the binomial distributions described in Lemma 3.1. For Lemma 3.2 this induction shows that w.h.p. the sequence $d(u), d_2(u), \ldots$ is bounded above by the sequence $a_1 np, a_2(np)^2, \ldots$, where the a_i satisfy a recurrence relation. This recurrence can later be solved to give bounds on the sequence a_i based on the exceptional probability desired. This strategy is inspired by [8].

Lemma 3.2. *Let $\mathcal{G} \sim \mathcal{G}(n, p)$, where $np = \omega(1)$. Then for $u \in V$ and $\alpha \in \mathbb{R}$, $\alpha \geq 6$,*

$$\mathbb{P}(|B_i(u)| > \alpha(np)^i) = o(e^{-\alpha np/3}).$$

Proof. Let $\mathcal{E}_i := \{d_i(u) \leq a_i(np)^i\}$ and $\mathcal{H}_i := \bigcap_{j=0}^{i} \mathcal{E}_i$. We shall show

$$\mathbb{P}\left(\mathcal{H}_i^c\right) \leq \sum_{j=0}^{i} e^{-j} \exp\left(-\frac{\lambda^2}{2(1 + \lambda/3\sqrt{np})}\right),$$

by induction on $i \geq 0$, where $a_i \geq 0$ is given by the recurrence

$$a_{i+1} = a_i + \frac{\lambda_i \sqrt{a_i}}{(np)^{(i+1)/2}}, \ a_0 = 1,$$

and $\lambda_i = \sqrt{3i + \lambda^2}$ for some λ specified later. For the base case $d_0(u) = 1 = a_0$. Now observe

$$a_i(np)^i np + \lambda_i \sqrt{a_i(np)^{i+1}} = (np)^{i+1}\left(a_i + \lambda_i \frac{\sqrt{a_i}}{(np)^{(i+1)/2}}\right) = a_{i+1}(np)^{i+1}. \tag{26}$$

Conditional on $d_i(u)$ we have $d_{i+1}(u) \preccurlyeq_1 Bin(d_i(u) \cdot n, p)$. Thus by (26) above

$$\begin{aligned}\mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) &:= \mathbb{P}(\{d_{i+1}(u) > a_{i+1}(np)^{i+1}\} \cap \mathcal{H}_i)\\ &\leq \mathbb{E}[\mathbb{P}(Bin(d_i(u) \cdot n, p) > a_i(np)^i np + \lambda_i \sqrt{a_i(np)^{i+1}} \mid d_i(u))\mathbf{1}_{\mathcal{H}_i}].\end{aligned}$$

Now by the Chernoff bounds, Lemma A.1, we have

$$\begin{aligned}\mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) &\leq \mathbb{E}\left[\exp\left(-\frac{\lambda_i^2 a_i(np)^{i+1}}{2(a_i(np)^{i+1} + \lambda_i\sqrt{a_i(np)^{i+1}}/3)}\right)\mathbf{1}_{\mathcal{H}_i}\right]\\ &= \mathbb{E}\left[\exp\left(-\frac{3i + \lambda^2}{2(1 + \lambda_i/3\sqrt{a_i(np)^{i+1}})}\right)\mathbf{1}_{\mathcal{H}_i}\right] \leq e^{-i}e^{-\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})}}\,\mathbb{P}(\mathcal{H}_i),\end{aligned} \tag{27}$$

for n large enough since $a_i \geq 1$ and $np = \omega(1)$ thus $\lambda_i / \sqrt{a_i(np)^{i+1}} \leq \lambda / \sqrt{np}$. Now observe that $\mathcal{H}_{i+1} \subseteq \mathcal{H}_i$ and $\mathcal{H}_i$ is the disjoint union of $\mathcal{H}_{i+1}$ and $(\mathcal{E}_{i+1})^c \cap \mathcal{H}_i$. Hence by (27) we have

$$\mathbb{P}(\mathcal{H}_{i+1}) = \mathbb{P}(\mathcal{H}_i) - \mathbb{P}((\mathcal{E}_{i+1})^c \cap \mathcal{H}_i) = \left(1 - e^{-i}\exp\left(-\frac{\lambda^2}{2(1 + \lambda/3\sqrt{np})}\right)\right)\mathbb{P}(\mathcal{H}_i).$$

If we continue iteratively and recall $\mathcal{H}_0 = \{d_0(u) \leq 1\}$, thus $\mathbb{P}(\mathcal{H}_0) = 1$, then we have

$$\mathbb{P}(\mathcal{H}_{i+1}) = \prod_{j=0}^{i}\left(1 - e^{-j} \cdot e^{-\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})}}\right)\mathbb{P}(\mathcal{H}_0) \geq 1 - \sum_{j=0}^{i} e^{-j} \cdot e^{-\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})}}.$$

Let $\lambda = k\sqrt{np}$ for any $k \geq 3$ and observe that

$$\mathbb{P}((\mathcal{H}_i)^c) \leq \sum_{j=0}^{i} e^{-j} \cdot e^{-\frac{\lambda^2}{2(1+\lambda/3\sqrt{np})}} = O\left(\exp\left(-\frac{k^2 np}{2(1 + k/3)}\right)\right) = O(e^{-3knp/4}), \tag{28}$$

where the last equality follows since $\frac{k^2}{2(1+k/3)} \geq \frac{3k}{4}$ provided $k \geq 3$, and $np = \omega(1)$.

We will show that $a_i \leq 2k$ for all i. Since $a_0 = 1 \leq 2k$ assume $a_i \leq 2k$, then by (26)

$$a_{i+1} = a_i + \frac{\lambda_i\sqrt{a_i}}{(np)^{(i+1)/2}} = 1 + \frac{\lambda_0\sqrt{a_0}}{\sqrt{np}} + \sum_{j=1}^{i}\frac{\lambda_j\sqrt{a_j}}{(np)^{(j+1)/2}}.$$

Recall that $\lambda_i = \sqrt{3i + \lambda^2}$ and observe that $\lambda_0 = \lambda = k\sqrt{np}$. Thus we have

$$a_{i+1} = 1 + k + \sum_{j=1}^{i}\frac{\sqrt{3j + k^2 np}\,\sqrt{2k}}{(np)^{(j+1)/2}} = 1 + k + O((np)^{-1/2}) \leq 2k,$$

for large n. Finally, conditional on $\bigcap_{j=0}^{i}\{d_j(u) \le 2k(np)^i\} \subseteq \mathcal{H}_i$ we have

$$|B_i(u)| = \sum_{j=0}^{i} d_j(u) \le \sum_{j=0}^{i} 2k(np)^i \le \alpha(np)^i,$$

where $\alpha = (2 + 1/100)k$, now for this α we have $\mathbb{P}((\mathcal{H}_i)^c) = o(e^{-\alpha np/3})$ by (28). □

Lemma 3.3. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, and $i \in \mathbb{Z}$ satisfy $1 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3$. Let Ψ_2 be defined with respect to* MBFS$(\mathcal{G}, \{u, v\})$ *for $u, v \in V$ such that $u \ne v$.*

(i) *Let $c > 0$, $np \ge c \log n$. Then $\mathbb{P}\left(d_i^*(y) < 15(np)^{i-1} \mid y \in \Psi_2\right) = o(e^{-4np})$.*
(ii) *If $np = \omega(\log n)$ then $\mathbb{P}\left(d_i^*(y) < \frac{9}{10}(np)^i \mid y \in \Psi_2\right) = o(n^{-K})$ for any fixed $K \ge 0$.*
(iii) *If $np - \log n \to \infty$ then $\mathbb{P}_{\mathcal{C}}(|B_j(v)| < 15(np)^{i-5}) = o(n^{-4})$ for any $5 \le j \le i-2$.*

Proof. We will first set up the general framework for a neighbourhood growth bound and then apply this bound under different conditions to prove Items (i), (ii), and (iii).

Run MBFS$(\mathcal{G}, \{u, v\})$ and let $y \in I_h$, $n_i := |S_{i+h}|$, $p_i := p \cdot (1-p)^{|I_{i+h}|-1}$, and $r_i = \prod_{j=i_0}^{i} n_j p_j$. We wish to show that there exists some $i_0 \in \mathbb{Z}$, $i_0 \ge 0$ such that for all $i \ge i_0$:

$$\mathbb{P}\left(d_{i+1}^*(y) < a_{i+1} r_i\right) \le (i+1)\exp(-\lambda^2/2), \tag{29}$$

where $a_i \ge 0$ satisfies $a_{i+1} = a_i - \lambda\sqrt{a_i}/\sqrt{r_i}$, for some initial a_{i_0} we will find later. Observe

$$a_i r_{i-1} n_i p_i - \lambda\sqrt{a_i r_{i-1} n_i p_i} = \left(a_i - \lambda\frac{\sqrt{a_i}}{\sqrt{r_i}}\right) r_i = a_{i+1} r_i.$$

Applying Lemma 3.1 (iii) and conditioning on $\mathfrak{F}_{i+h}$ yields $d_{i+1}^*(y) \sim Bin(n_i, d_i^*(y)p_i)$.

Let $\mathcal{H}_i := \left\{d_i^*(u) \ge a_i r_{i-1}\right\} \in \mathfrak{F}_{i+h}$ and assume $\mathbb{P}(\mathcal{H}_i^c) \le ie^{-\lambda^2/2}$. Now by Lemma A.1 (i),

$$\begin{aligned}
\mathbb{P}\left(d_{i+1}^*(y) < a_{i+1} r_i\right) &= \mathbb{E}\left[\mathbb{P}\left(d_{i+1}^*(y) < a_i r_{i-1} n_i p_i - \lambda\sqrt{a_i r_{i-1} n_i p_i} \mid \mathfrak{F}_{i+h}\right)\right] \\
&\le \mathbb{E}\left[\mathbb{P}\left(Bin\left(n_i, d_i^*(y)p_i\right) < a_i r_i - \lambda\sqrt{a_i r_i} \mid \mathfrak{F}_{i+h}\right)\mathbf{1}_{\mathcal{H}_i}\right] + \mathbb{P}\left(\mathcal{H}_i^c\right) \\
&\le \exp(-\lambda^2 a_i r_i/(2a_i r_i)) + i\exp(-\lambda^2/2) = (i+1)\exp(-\lambda^2/2).
\end{aligned}$$

The above always holds, however, it may be vacuous as if i is too large then a_i may be negative. This can also happen for an incorrect choice of the starting time i_0 and initial value a_{i_0}. We address this in the application making sure to condition on events where everything is well defined. In this spirit let $\ell := \lfloor \log(n)/\log(np) \rfloor - h - 1$ and define the event

$$\mathcal{I} := \bigcap_{i=0}^{\ell} \{|I_{i+h}| \le 26(np)^{i+h}\} \cap \left\{d_i^*(y) \le 13(np)^i\right\} \cap \{|S_{i+h}| \ge n - 26(np)^{i+h}\}.$$

Conditioning on the event $\mathcal{I}$ and the filtration $\mathfrak{F}_{i+h}$ for any $i \le \ell$ ensures that $Bin(n_i, d_i^*(y)p_i)$ is a valid probability distribution and $n_i p_i = (1 - o(1))np$. By Lemma 3.2 with $\alpha = 13$,

$$\mathbb{P}(\mathcal{I}^c) \le 2\sum_{i=0}^{\ell+2} \mathbb{P}(|B_i(u)| > 13(np)^i) = o(e^{-4np}). \tag{30}$$

Item (i): Recall from (23) that if $y \in \Psi_2(u) \cup \Psi_2(v) \subseteq I_2$ then $d^*(y) > D$, defined at (21). Thus conditional on $\mathcal{I} \cap \mathfrak{F}_3$, $d_2^*(y) \succcurlyeq_1 Bin(n(1 - o(1)), Dp(1 - o(1)))$. If we choose $\lambda = 3\sqrt{np}$, then since in this regime $D = \max\{\lceil \frac{50}{c} \rceil, 50\}$ applying Lemma A.1 (i) yields

$$\mathbb{P}\Big(d_2^*(y) < Dn_1p_1/2\Big) = \mathbb{E}[\mathbb{P}\Big(d_2^*(y) < Dn_1p_1/2 | \mathfrak{F}_3\Big)(\mathbf{1}_{\mathcal{I}} + \mathbf{1}_{\mathcal{I}^c})] \le e^{-Dnp/10} + \mathbb{P}(\mathcal{I}^c) \le e^{-\lambda^2/2}.$$

Take $i_0 = 1$ and $a_2 = D/3$ since on $\mathcal{I}$ we have $D/2n_1p_1 \ge Dnp/3$. Now $a_2 \ge \cdots \ge a_i$ so on the event $\mathcal{I}$ we have the following for any $3 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3$:

$$a_i = a_2 - \sum_{k=2}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \ge \frac{D}{3} - (3 + o(1))\sqrt{\frac{D}{3np}} \ge 16,$$

since conditional on $\mathcal{I}$ we have $r_i = \prod_{j=i_0}^{i} n_j p_j \ge (1 - o(1))(np)^i$ for any $1 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3$. Notice also that $d^*(y) > D > 15(np)^0$ so by (29)

$$\mathbb{P}\Big(d_{i+1}^*(y) < 15(np)^i\Big) \le \mathbb{P}\Big(d_{i+1}^*(y) < a_{i+1}r_i\Big) + \mathbb{P}(\mathcal{I}^c) \le (i+1)e^{-\lambda^2/2} + o(e^{-4np}) = o(e^{-4np}).$$

Item (ii): For this case on the event $\mathcal{I}$ we have $n_i p_i = (1 - o(1))np = \omega(\log n)$ for every $0 \le i \le \ell$, this is why $D = 0$ when $np = \omega(\log n)$ in the definition (21) as we do not need to rely on the fact that $d^*(y)$ greater than some constant to start the branching.

Fix $K > 0$ and let $\lambda = \sqrt{3K \log n}$. As before conditioning on $\mathcal{I} \cap \mathfrak{F}_3$ ensures that $d^*(y) \sim Bin(n_0, p_0) \succcurlyeq_1 Bin(n(1 - o(1)), p(1 - o(1)))$. By Lemma A.1 (i),

$$\mathbb{E}[\mathbb{P}(d^*(y) < r_0 - (5/4)\lambda\sqrt{r_0} | \mathfrak{F}_3)(\mathbf{1}_{\mathcal{I}} + \mathbf{1}_{\mathcal{I}^c})] \le e^{-25\lambda^2/32} + \mathbb{P}(\mathcal{I}^c) \le \exp(-\lambda^2/2).$$

Take $i_0 = 0$, $a_1 = 19/20$ since on $\mathcal{I}$ we have $r_0 - (5/4)\lambda\sqrt{r_0} \ge 19np/20$. Now $a_1 \ge \cdots \ge a_i$ so on the event $\mathcal{I}$ we have the following for any $2 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3$

$$a_i = a_1 - \sum_{k=1}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \ge \frac{19}{20} - (1 + o(1))\frac{\sqrt{19 \cdot 3K \log n}}{\sqrt{20np}} = \frac{19}{20} - o(1) \ge \frac{9}{10}.$$

Thus for any $1 \le i \le \lfloor \log(n)/\log(np) \rfloor - 3$, $K > 0$ we have

$$\mathbb{P}\Big(d_i^*(y) < 9/10(np)^i\Big) \le \mathbb{P}\Big(d_i^*(y) < a_i r_{i-1}\Big) + \mathbb{P}(\mathcal{I}^c) \le (i+1)e^{-\lambda^2/2} + e^{-4np} \le o(n^{-K}).$$

Item (iii): We assume that $np = \theta(\log n)$ for larger p the result follows from stochastic domination. Since $\mathcal{G} \in \mathcal{C}$ there exists a path $u := u_0, u_1, \ldots, u_4$ with $u_{j-1}u_j \in E$ for each $1 \le j \le 4$. Let $f(u_j) = |\{v \in V \backslash \{u_0, \ldots, u_4\} : u_j v \in E\}|$. Then, for D from (21), by Lemma A.1 (ii),

$$\mathbb{P}(f(u_j) < D) = \mathbb{P}(Bin(n-\ell-1,p) < D) \le e^{-(n-\ell-1)p} \cdot \left(\frac{e(n-\ell-1)p}{D}\right)^D = e^{-(1-o(1))np}.$$

Let $\mathcal{E}$ be the event $\{d(u_{j_0}) \ge D$ for some $0 \le j_0 \le 4\}$. Then as $\{f(u_j)\}_{j=0}^4$ are i.i.d. we have

$$\mathbb{P}_{\mathcal{C}}(\mathcal{E}^c) \le \mathbb{P}(f(u_j) < D)^5/\mathbb{P}(\mathcal{C}) \le e^{-5(1-o(1))\log n} \le o(n^{-4}). \tag{31}$$

On $\mathcal{E}$ there is some $u_{j_0} \in V$ with $d(u,u_{j_0}) = j_0 \le 4$ and $d(u_{j_0}) > D$. We use the stochastic domination $d_i(u_{j_0}) \succcurlyeq_1 d_i^*(u_{j_0})$ to bound the growth of $|B_{i+j_0}(u)|$ from below by that of $d_i^*(u_{j_0})$, where $u_{j_0} \in I_{j_0}$ is defined with respect to MBFS$(\mathcal{G},\{u,v\})$ for some $v \in V$. Let $\lambda = 3\sqrt{\log n}$ and recall $D = \max\{\lceil\frac{50}{c}\rceil, 50\}$. On $\mathcal{I}$, $r_{j_0+1} = (1-o(1))np$, thus by Lemma A.1 (i):

$$\mathbb{E}\left[\mathbb{P}\left(d^*_{j_0+2}(u) < dn_{j_0+1}p_{j_0+1}/2 | \mathfrak{F}_{j_0+1}\right)\mathbf{1}_{\mathcal{I}\cap\mathcal{E}}\right] \le \mathbb{E}[e^{-dr_{j_0+1}/8}\mathbf{1}_{\mathcal{I}\cap\mathcal{E}}] \le e^{-\lambda^2/2}.$$

Take $i_0 = j_0 + 1$ and $a_{j_0+2} = d/3$ since on $\mathcal{I}\cap\mathcal{E}$ we have $dn_{j_0+1}p_{j_0+1}/2 \ge dnp/3$. Now $a_{i_0} \ge \cdots \ge a_i$ and on the event $\mathcal{I}\cap\mathcal{E}$ we have $r_i = (1-o(1))(np)^{i-j_0}$. Thus we have the following for any $\varepsilon > 0$ and $j_0 + 3 \le i \le \lfloor \log(n)/\log(np)\rfloor - j_0 - 1$:

$$a_i = a_{j_0+2} - \sum_{k=j_0+2}^{i-1} \frac{\lambda\sqrt{a_k}}{\sqrt{r_k}} \ge \frac{d}{3} - (3+o(1))\sqrt{\frac{d\log n}{3(np)^2}} \ge 16.$$

Notice also $d^*_{j_0+1}(y) > d > 15(np)^0$. Thus for any $4 \le i \le \lfloor \log(n)/\log(np)\rfloor - 5$:

$$\begin{aligned}\mathbb{P}_{\mathcal{C}}(|B_{i+1}(u)| < 15(np)^{i-5}) &\le \mathbb{P}_{\mathcal{C}}(d_{i+1}(u) < 15(np)^{i-j_0-1} | \mathcal{I},\mathcal{E}) + \mathbb{P}_{\mathcal{C}}(\mathcal{I}^c) + \mathbb{P}_{\mathcal{C}}(\mathcal{E}^c)\\ &\le \mathbb{P}\left(d^*_{i+1}(y) < a_{i+1}r_i | \mathcal{I},\mathcal{E}\right)/\mathbb{P}(\mathcal{C}) + \mathbb{P}(\mathcal{I}^c)/\mathbb{P}(\mathcal{C}) + \mathbb{P}_{\mathcal{C}}(\mathcal{E}^c),\end{aligned}$$

which is at most $2(i+1)e^{-\lambda^2/2} + o(e^{-4np}) + o(n^{-4}) = o(n^{-4})$ by the bounds on $\mathbb{P}(\mathcal{C})$, $\mathbb{P}(\mathcal{I})$, and $\mathbb{P}(\mathcal{E}^c)$ given by (10), (30), and (31), respectively. □

The next lemma allows us to couple the complex Ψ_1 and Φ neighbourhood distributions to the far more simple Γ^* and Γ-neighbourhood distributions.

Lemma 3.4. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, where $c\log n \le np \le o(n^{1/3})$ for any $c > 0$. Let I_1 and the φ, ψ, ψ_2, and d^*-distributions be defined with respect to* MBFS$(\mathcal{G},\{u,v\})$, *$u,v \in V$. Then*

(i) $\mathbb{P}(\varphi(x) \ne d^*(x) | x \in I_1) = e^{-(1-o(1))np}$,
(ii) $\mathbb{P}(\psi(u) \ne d^*(u)) = e^{-(1-o(1))np}$,
(iii) $\mathbb{P}(\psi(u) \ne d(u)$ or $\psi(v) \ne d(v)) \le 2np^2 + e^{-(1-o(1))np}$,
(iv) $\mathbb{P}(\psi_2(u) \ne d_2^*(u)$ or $\psi_2(v) \ne d_2^*(v)) = e^{-(1-o(1))np}$,
(v) $\mathbb{P}(\psi_2(u) \ne d_2(u)$ or $\psi_2(v) \ne d_2(v)) \le 4n^3p^4 + e^{-(1-o(1))np} + O(n^2p^3)$.

Proof.
Item (i): Run MBFS($\mathcal{G}, \{u, v\}$) and let $x \in I_1$. By the definition (20) of $\psi(x)$, if $d^*(\tilde{x}) > D$ for all $\tilde{x} \in \Gamma_1^*(x)$ then $\varphi(x) = d^*(x)$, where D is bounded (21). Hence for $x \in I_1$,

$$\mathbb{P}(\varphi(x) \neq d^*(x)|\mathfrak{F}_2) = \mathbb{P}\Big(d^*(\tilde{x}) \leq D \text{ for some } \tilde{x} \in \Gamma_1^*(x)|\mathfrak{F}_2\Big) \leq \sum_{\tilde{x}\in\Gamma_1^*(x)} \mathbb{P}(d^*(\tilde{x}) \leq D|\mathfrak{F}_2). \tag{32}$$

If $\tilde{x} \in \Gamma_1^*(x), x \in I_1$ then $\tilde{x} \in I_2$. Knowing the parent of $\tilde{x}$ does not affect the d^*-distribution conditioned on $\{\tilde{x} \in I_2\}$, so by Lemma 3.1 (iii) as $|S_2|, |I_2| \in \mathfrak{F}_2$ we have

$$\mathbb{P}(\varphi(x) \neq d^*(x)|\mathfrak{F}_2) \leq d^*(x)\mathbb{P}(Bin(|S_2|, p(1-p)^{|I_2|-1}) \leq D||S_2|, |I_2|).$$

Let $\mathcal{E}_x^j := \{|I_j| \leq 12(np)^j\} \cap \{d^*(x) \leq 6np\} \cap \{|S_j| \geq n - 12(np)^j\}$ for $x \in I_{j-1}$ and observe

$$\mathbb{P}\left(\left(\mathcal{E}_x^j\right)^c\right) \leq 2\mathbb{P}(|B_j(u)| > 6(np)^j) + \mathbb{P}(|B_1(x)| > 6np) = o(e^{-2np}), \tag{33}$$

by Lemma 3.2 with $\alpha = 6$. Now by Lemma A.1 (ii) and Bernoulli's inequality (A1)

$$\mathbb{P}(\varphi(x) = d^*(x)|\mathfrak{F}_2)\mathbf{1}_{\mathcal{E}_x^1} \leq d^*(x)e^{-|S_2|p(1-p)^{|I_2|}}\left(\frac{e|S_2|\,p(1-p)^{|I_2|}}{D}\right)^D\mathbf{1}_{\mathcal{E}_x^1} \leq e^{-(1-o(1))np},$$

the result follows by taking expectation of the above and using (33) with $j = 1$.
Item (ii): for $\tilde{u} \in I_1$ the distribution of $d^*(\tilde{u})$ conditioned on $|S_1|, |I_1|$ is known by 3.1 (iii). Thus using the bound $(1-p)^n \leq \exp(-np)$ we obtain the following for $\tilde{u} \in I_1$,

$$\mathbb{P}(d^*(\tilde{u}) = 0|\mathfrak{F}_1) = \mathbb{P}(Bin(|S_1|, p(1-p)^{|I_1|-1}) = 0|\mathfrak{F}_1) \leq \exp(-|S_1|\,p(1-p)^{|I_1|}). \tag{34}$$

Recall the definition (22) of $\Psi_1(u)$. If $\tilde{u} \in \Gamma_1^*(u)$ then $\tilde{u} \in I_1$ and knowing the parent of $\tilde{u}$ does not affect the d^*-distribution conditioned $\{\tilde{u} \in I_1\}$. Thus, similarly to (32), for $\tilde{u} \in \Gamma_1^*(u)$,

$$\mathbb{P}(\psi(u) \neq d^*(u)|\mathfrak{F}_1) = \mathbb{P}(\varphi(\tilde{u}) = 0 \text{ for some } \tilde{u} \in \Gamma_1^*(u)|\mathfrak{F}_1) \leq d^*(u) \cdot \mathbb{P}(\varphi(\tilde{u}) = 0|\mathfrak{F}_1).$$

Now using the coupling inequality (A3), yields the following for $\tilde{u} \in I_1$:

$$\mathbb{P}(\psi(u) \neq d^*(u)|\mathfrak{F}_1) \leq d^*(u)\mathbb{P}(d^*(\tilde{u}) = 0|\mathfrak{F}_1) + d^*(u)\mathbb{P}(\varphi(\tilde{u}) \neq d^*(\tilde{u})|\mathfrak{F}_1). \tag{35}$$

Recall $\mathcal{E}_u^1$ from (33) and observe $\mathbb{P}(d^*(\tilde{u}) = 0|\mathfrak{F}_1)\mathbf{1}_{\mathcal{E}_x^1} \leq e^{-(1-o(1))np}$ by (34). Thus

$$\mathbb{P}(\psi(u) \neq d^*(u)) \leq e^{-(1-o(1))np} + 6np\mathbb{P}(\varphi(\tilde{u}) \neq d^*(\tilde{u})|\tilde{u} \in I_1) + o(e^{-2np}) = e^{-(1-o(1))np},$$

by taking the expectation of (35) conditioned on $\mathcal{E}_u^1$ then using the bound on $\mathbb{P}(d^*(\tilde{u}) = 0|\mathfrak{F}_1)$ above and bounds on $\mathbb{P}(\varphi(\tilde{u}) \neq d^*(\tilde{u})|\tilde{u} \in I_1)$ and $\mathbb{P}((\mathcal{E}_u^1)^c)$ from Item (i) and (33) respectively.

Item (iii): let $I_0 = \{u, v\}$ and $\mathcal{H} := \{d(u) = d^*(u),\ d(v) = d^*(v)\}$. By Item (ii)

$$\begin{aligned}\mathbb{P}(\psi(u) \neq d(u) \text{ or } \psi(v) \neq d(v)) &\leq \mathbb{P}(\{\psi(u) \neq d(u) \text{ or } \psi(v) \neq d(v)\} \cap \mathcal{H}) + \mathbb{P}(\mathcal{H}^c)\\ &\leq \mathbb{P}(\psi(u) \neq d^*(u) \text{or} \psi(v) \neq d^*(u)) + \mathbb{P}(\mathcal{H}^c)\\ &\leq 2e^{-(1-o(1))np} + \mathbb{P}(\mathcal{H}^c).\end{aligned} \tag{36}$$

To calculate $\mathbb{P}(\mathcal{H}^c)$ in the above recall the definition (17) of $d^*(u)$ and observe

$$\begin{aligned}\mathbb{P}(\mathcal{H}^c) &= \mathbb{P}(\{uv \in E\} \cup \{xu \in E \text{ and} xv \in E \text{ for some } x \in V \backslash I_0\})\\ &\leq \mathbb{P}(uv \in E) + \sum_{x \in V \backslash I_0} \mathbb{P}(xu \in E \text{ and} xv \in E) = p + (n-2)p^2.\end{aligned} \tag{37}$$

Finally, combining (36) and (37) yields the bound

$$\mathbb{P}(\psi(u) \neq d(u) \text{ or } \psi(v) \neq d(v)) \leq 2e^{-(1-o(1))np} + p + (n-2)p^2 \leq 2np^2 + e^{-(1-o(1))np}.$$

Item (iv): Define the following events

$$\mathcal{J} := \bigcap_{x \in I_1} \{\varphi(x) = d^*(x)\}, \quad \mathcal{D} = \{d(u), d(v) \leq (1 + 9/\min\{c, 1\})np\}.$$

Notice $\{\psi_2(u) \neq d_2^*(u) \text{or} \psi_2(v) \neq d_2^*(u)\} = \{\psi(u) \neq d^*(u) \text{ or } \psi(v) \neq d^*(u)\} \cup \mathcal{J}^c$. Now

$$\mathbb{P}(\mathcal{J}^c) \leq \mathbb{E}[|I_1| \, \mathbb{P}(\varphi(x) = d^*(x) | \mathfrak{F}_1)(\mathbf{1}_{\mathcal{D}} + \mathbf{1}_{\mathcal{D}^c})] \leq O\left(npe^{-(1-o(1))np}\right) + n\mathbb{P}(\mathcal{D}^c), \tag{38}$$

which is $o(e^{-np(1-o(1))})$ by Item (i) and since $\mathbb{P}(\mathcal{D}^c) = o(e^{-np}/n)$ by Lemma A.1 (i).

Item (v): Let $I_0 = \{u, v\}$ and $\mathcal{L} := \{d_2(u) = d_2^*(u), \ d_2(v) = d_2^*(v)\}$. Then

$$\mathcal{L} := \left(\bigcap_{x \in d(u), y \in d(v)} \{xy \notin E\} \right) \cap \left(\bigcap_{z \in S_1} \{|\{x \in I_1 : xz \in E\}| \leq 1\} \right) \cap \mathcal{H}, \tag{39}$$

by the definition (17) of $d_2^*(u)$. Observe that by the Bernoulli inequality (A1),

$$\begin{aligned}&\mathbb{P}(|\{x \in I_1 : xz \in E\}| > 1 | \mathfrak{F}_1) = 1 - \sum_{a=0,1} \mathbb{P}(|\{x \in I_1 : xz \in E\}| = a | \mathfrak{F}_1)\\ &= 1 - (1-p)^{|I_1|} - |I_1| \, p(1-p)^{|I_1|-1} \leq 1 - (1 - |I_1| \, p) - |I_1| \, p(1 - |I_1| \, p) = (|I_1| \, p)^2.\end{aligned}$$

By (39), the above estimate on $\mathbb{P}(|\{x \in I_1 : xz \in E\}| > 1 | \mathfrak{F}_1)$ and $\mathcal{H} \in \mathfrak{F}_1$, we have

$$\begin{aligned}\mathbb{P}(\mathcal{L}^c | \mathfrak{F}_1) &\leq \sum_{x \in d(u), y \in d(v)} \mathbb{P}(xy \in E | \mathfrak{F}_1) + \sum_{z \in S_1} \mathbb{P}(|\{x \in I_1 : xz \in E\}| > 1 | \mathfrak{F}_1) + \mathbb{P}(\mathcal{H}^c | \mathfrak{F}_1)\\ &\leq d(u)d(v)p + |S_1|(|I_1| \, p)^2 + \mathbb{P}(\mathcal{H}^c | \mathfrak{F}_1).\end{aligned}$$

Then by the bound on $\mathbb{P}(\mathcal{L}^c | \mathfrak{F}_1)$ above, the tower property and Cauchy Schwartz inequality,

$$\mathbb{P}(\mathcal{L}^c) \leq p\sqrt{\mathbb{E}[d(u)^2]\mathbb{E}[d(v)^2]} + p^2\sqrt{\mathbb{E}[|S_1|^2]\mathbb{E}[|I_1|^4]} + \mathbb{P}(\mathcal{H}^c) = 4n^3p^4 + O\left(n^2p^3\right), \tag{40}$$

the last equality holds by (37) and as $|S_1| \sim Bin(n-2,(1-p)^2)$, $|I_1| \sim Bin(n-2, 2p(1-p))$ by Lemma 3.1, then applying the bound on moments of binomial r.v.'s from (A4).

Finally let $\mathcal{F} := \{\psi_2(u) = d_2(u),\ \psi_2(v) = d_2(v)\}$. Then by the definitions (22) and (23) of the vertex sets $\Psi_1(u)$ and $\Psi_2(u)$ we have $\mathcal{F} :=\{\psi(u) = d(u),\ \psi(v) = d(v)\} \cap \mathcal{J} \cap \mathcal{L}$. Thus

$$\mathbb{P}(\mathcal{F}^c) \le \mathbb{P}(\psi(u) \ne d(u) \text{ or } \psi(v) \ne d(u)) + \mathbb{P}(\mathcal{J}^c) + \mathbb{P}(\mathcal{L}^c) \le 4n^3p^4 + e^{-(1-o(1))np} + O\left(n^2p^3\right),$$

by Item (iii), (40) and (38). □

Recall from Definition (24) that $\mathcal{A}_{u,v}$ is the event that $\mathcal{G}(n,p)$ satisfies the strong path property for $u, v \in V$ and some $k \le \log(np)/(2\log n) + 2$. Recall also Definition (25) of $\mathcal{B}_{u,v}$ which is the event the pruned first neighbourhoods $\Psi_1(u), \Psi_1(v)$ are both nonempty.

Lemma 3.5. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, where $c\ n \le np < n^{1/10}$, $c > 0$. Then for $u, v \in V,\ \ u \ne v$,*

$$\mathbb{P}((\mathcal{A}_{u,v})^c) = o(e^{-7\min\{np,\log n\}/2}) \quad \text{and} \quad \mathbb{P}((\mathcal{B}_{u,v})^c) = e^{-(1-o(1))np}.$$

Proof. Run MBFS($\mathcal{G}, \{u,v\}$), $u, v \in V$. For $k \ge 0$ let $\mathcal{T} := \mathcal{T}_{u,v,k} = \mathcal{T}_1 \cap \mathcal{T}_2$, where

$$\mathcal{T}_1 := \{|S_{k+2}| \ge n - n^{5/6}\}, \quad \mathcal{T}_2 := \left\{|\Gamma_k^*(x) \times \Gamma_k^*(y)| \ge 4n,\ \text{for all } (x,y) \in \Psi_2(u) \times \Psi_2(v)\right\}.$$

On the event $\mathcal{T}_1$ when MBFS($\mathcal{G}, \{u,v\}$) has run for $k+2$ iterations there is still a lot of the graph yet to explore and the algorithm will run for at least one more iteration. The k in the definition of $\mathcal{T}$ will be the one occurring in the description of $\mathcal{A}_{u,v}$. Set the value of k to be

$$k := k(n,p) = \begin{cases} \left\lceil \log\left(\dfrac{4n}{(15)^2}\right)/(2\log(np)) \right\rceil + 1, & \text{if } np = c\log n \text{ where } c > 0, \\ \left\lceil \log\left(\dfrac{400n}{81}\right)/(2\log(np)) \right\rceil, & \text{if } np = \omega(\log n). \end{cases} \tag{41}$$

Thus $k \le \log(np)/(2\log n) + 2$ for large n. It remains to show that for k given by (41):

$$\mathbb{P}\left(\mathcal{G} \notin A_{u,v}^{n,k}\right) = o\left(e^{-7\min\{np,\ \log n\}/2}\right).$$

Let $\mathcal{R} := \mathcal{R}_{u,v}$ be the event $\{|\Psi_2(u) \times \Psi_2(v)| \le (12(np)^2)^2\}$. Since $\psi_2(u) \le d_2(u)$ for any $u \in V$, we have $\mathbb{P}(\mathcal{R}^c) = o(e^{-4np})$ by Lemma 3.2 with $\alpha = 12$. Thus by the tower property

$$\mathbb{P}(\mathcal{T}^c) \le \mathbb{P}\left(\mathcal{T}_1^c\right) + \mathbb{P}\left(\mathcal{R}^c\right) + \mathbb{E}\left[\mathbb{P}\left(\mathcal{T}_2^c \middle| \mathfrak{F}_3\right)\mathbf{1}_{\mathcal{R}}\right] \le 2\mathbb{P}\left(|B_{k+2}(u)| > n^{5/6}/2\right) + 2\mathbb{P}\left(|B_2(u)| > 12(np)^2\right)$$
$$+2\mathbb{E}\left[\psi_2(u)\psi_2(v)\mathbf{1}_{\mathcal{R}}\mathbb{P}\left(d_k^*(w) < 2n^{1/2} \middle| \{w \in \Psi_2\}, \mathfrak{F}_3\right)\right],$$

as $\{d^*(x)\cdot d^*(y) < k\} \subseteq \{d^*(x) \text{ or } d^*(y) < \sqrt{k}\}$. Provided $np \le n^{1/10}$ the choice of k given by (41) satisfies the conditions of Lemma 3.3. Thus by Lemmas 3.2, 3.3 (i), and 3.3 (ii) we have

$$\mathbb{P}(\mathcal{T}^c) \le o\left(e^{-9np/2}\right) + O(np)^4 \cdot \mathbb{P}\left(d_k^*(w) < 2n^{1/2} \mid w \in \Psi_2\right) = o\left(e^{-7\min\{np,\log n\}/2}\right), \tag{42}$$

where the bound $\mathbb{P}(d_k^*(w) < 2n^{1/2} \mid w \in \Psi_2) \le e^{-4\min\{np,\log n\}}$ above covers the different values of np and comes from amalgamating Lemmas 3.3 (i) and (ii), with $K = 4$ in the latter.

Let $\mathcal{L}_{x,y}$ be the following event indexed by $(x,y) \in \Psi_2(u) \times \Psi_2(v)$,

$$\mathcal{L}_{x,y} := \{x'y' \notin E, \text{ for every pair } (x',y') \in \Gamma_k^*(x) \times \Gamma_k^*(y)\}.$$

This is independent of $\mathfrak{F}_{k+2}$ as each $x'y'$ has not been checked up to iteration $k+2$, thus

$$\mathbb{P}(\mathcal{L}_{x,y} \mid \mathfrak{F}_{k+2})\mathbf{1}_{\mathcal{T}} = \mathbb{P}(x'y' \notin E)^{d_k^*(x)d_k^*(y)}\mathbf{1}_{\mathcal{T}} \le (1-p)^{4n} \le e^{-4np}. \tag{43}$$

Recall Definition 2.4 of the strong path property $A_{u,v}^{n,k}$ which we can express as

$$\left\{\mathcal{G} \notin A_{u,v}^{n,k}\right\} = \bigcup_{(x,y)\in\Psi_2(u)\times\Psi_2(v)} \left\{\Gamma_k^*(x) = \emptyset\right\} \cup \left\{\Gamma_k^*(y) = \emptyset\right\} \cup \mathcal{L}_{x,y}.$$

Observe that for each $i,j \ge 0$ the random variables $\{d_j^*(w)\}_{w\in I_i}$ are identically distributed. Recall also that $\Psi_1(u), \Psi_1(v), \mathcal{R} \in \mathfrak{F}_3$. Let $\mathfrak{P}$ denote $\mathbb{P}(\{\mathcal{G} \notin A_{u,v}^{n,k}\} \cap \mathcal{R} \cap \mathcal{T})$. By the union bound, tower property and since $\psi_1(u)\psi_1(v) \le 12^2(np)^4$ on $\mathcal{R}$, we have

$$\begin{aligned}\mathfrak{P} &\le \mathbb{E}\left[\sum_{(x,y)\in\Psi_2(u)\times\Psi_2(v)} \mathbb{E}[(\mathbf{1}_{\mathcal{L}_{x,y}\cup\{d_k^*(x)=0\}\cup\{d_k^*(y)=0\}})\mathbf{1}_{\mathcal{R}}\mathbf{1}_{\mathcal{T}} \mid \mathfrak{F}_3]\right] \\ &\le \mathbb{E}[12^2(np)^4\mathbf{1}_{\mathcal{R}}\mathbb{E}[(\mathbf{1}_{\mathcal{L}_{x,y}} + \mathbf{1}_{\{d_k^*(x)=0\}} + \mathbf{1}_{\{d_k^*(y)=0\}})\mathbf{1}_{\mathcal{T}} \mid \mathfrak{F}_3]].\end{aligned}$$

Now since $x,y \in \Psi_2$ and $d_j^*(x), d_j^*(y)$ are identically distributed for any $j \ge 0$:

$$\mathfrak{P} \le O(np)^4 \cdot \left(\mathbb{E}\left[\mathbb{E}\left[\mathbf{1}_{\mathcal{L}_{x,y}}\mathbf{1}_{\mathcal{T}} \,\middle|\, \mathfrak{F}_{k+2}\right]\right] + 2\mathbb{P}\left(d_k^*(w) = 0 \,\middle|\, \{w \in \Psi_2\}\right)\right).$$

By Lemma 3.3 (i), (43) and since $\mathcal{T} \in \mathfrak{F}_{k+2}$ we have

$$\mathfrak{P} \le O(np)^4 \cdot (\mathbb{E}[\mathbb{P}(\mathcal{L}_{x,y} \mid \mathfrak{F}_{k+2})\mathbf{1}_{\mathcal{T}}] + 2e^{-4\min\{np,\log n\}}) = o(e^{-7\min\{np,\log n\}/2}).$$

Recall $\mathbb{P}(\mathcal{R}^c) = o(e^{-4np})$, so by (42) and the bound on $\mathbb{P}(\{\mathcal{G} \notin A_{u,v}^{n,k}\} \cap \mathcal{R} \cap \mathcal{T})$ above,

$$\mathbb{P}\left(\mathcal{G} \notin A_{u,v}^{n,k}\right) \le \mathbb{P}\left(\left\{\mathcal{G} \notin A_{u,v}^{n,k}\right\} \cap \mathcal{R} \cap \mathcal{T}\right) + \mathbb{P}\left((\mathcal{R} \cap \mathcal{T})^c\right) \le o(e^{-7\min\{np,\log n\}/2}).$$

For $\mathbb{P}((\mathcal{B}_{u,v})^c)$ we apply the coupling inequality (A3) to the ψ and d^*-distributions:

$$\mathbb{P}((\mathcal{B}_{u,v})^c) \le \mathbb{P}(\psi_1(u) = 0) + \mathbb{P}(\psi_1(v) = 0) \le 2\mathbb{P}(d^*(u) = 0) + 2\mathbb{P}(\psi_1(u) \ne d^*(u)).$$

Then since $\mathbb{P}(\psi_1(u) \neq d^*(u))$ is known by Lemma 3.4 we have

$$\mathbb{P}((\mathcal{B}_{u,v})^c) \leq 2\mathbb{P}(d^*(u) = 0 \mid d^*(v) \leq 6np) + 2\mathbb{P}(d\ (v) > 6np) + 2e^{-(1-o(1))np}.$$
$$\leq 2(1-p)^{n-6np-1} + o(e^{-2np}) + e^{-(1-o(1))np} = e^{-(1-o(1))np},$$

by applying Lemma 3.1 (ii) to the first term and Lemma 3.2 (i) with $\alpha = 6$ to the second. □

The following crude but resilient bound on $R(i,j)$ is useful when conditioning on $\mathcal{A}_{i,j}^c$.

Lemma 3.6. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$ be such that $np - \log n \to \infty$. Then for $i, j \in V$,*

$$\mathbb{P}_{\mathcal{C}}(R(i,j) > 3\log n/\log(np)) = o(n^{-4}).$$

Proof. Since $G \in \mathcal{C}$ the effective resistance between two points is bounded from above by the graph distance. Let $\mathcal{J}_{i,j} := \{|B_k(i)| \cdot |B_k(j)| \geq 4n\}$, where $k := \lceil \log(\frac{4n}{15^2})/(2\log np) \rceil + 5$. Using Lemma 3.3 (iii) to bound $\mathbb{P}_{\mathcal{C}}(\mathcal{J}_{i,j}^c)$, since $5 \leq k \leq \lfloor \log(n)/\log(np) \rfloor - 5$ when n large:

$$\begin{aligned}
\mathbb{P}\mathcal{C}(R(i,j) > 2k+1) &\leq \mathbb{P}\mathcal{C}(d(i,j) > 2k+1 \mid \mathcal{J}_{i,j}) + \mathbb{P}\mathcal{C}\left(\mathcal{J}_{i,j}^c\right)\\
&\leq \mathbb{P}(xy \notin E,\ \forall\ (x,y) \in B_k(i) \times B_k(j), B_k(i) \cap B_k(j) = \emptyset \mid \mathcal{J}_{i,j})/\mathbb{P}(\mathcal{C})\\
&\quad + 2\mathbb{P}\mathcal{C}(|B_k(j)| < 2\sqrt{n}) \leq 2(1-p)^{4n} + 2\,o(n^{-4}) = o(n^{-4}).
\end{aligned}$$

The result follows since $2k+1 = 2(\lceil \log(\frac{4n}{15^2})/(2\log np) \rceil + 5) + 1 \leq \frac{3\log n}{\log(np)}$ for large n. □

4 | PROOF OF THEOREMS 1.1, 1.3, AND 1.6

Most of our main theorems result from Corollary 2.5, Lemma 4.1 below simplifies this application.

Lemma 4.1. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, where $\log n + \log\log\log n \leq np < o(n^{1/3})$. Let $\alpha \geq 1$ and $\Psi_1(u), \Psi_1(v)$ be defined with respect to* MBFS$(\mathcal{G}, \{u,v\})$, *$u, v \in V$. Then*

(i) $\mathbb{E}_{\mathcal{C}}[(\sup_{x \in \Psi_1(u)} \frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\varphi(x)})^\alpha]^{1/\alpha} = O\left(\frac{1}{np}\right)$.

(ii) *If $c \log n \leq np \leq n^{1/10}$, for any fixed $c > 0$, then*

$$\mathbb{P}\left(R(u,v) > \left(\frac{1}{\psi(u)} + \frac{1}{\psi(v)}\right)(1 + 9\varepsilon_n)\right) = o(e^{-np/4}) + o(n^{-7/2}).$$

Proof.
Item (i): Let $\mathfrak{P}_u := \mathbb{P}(\inf_{x \in \Psi_1(u)} \varphi(x) < K_p)$ and $K_p := (1 - \sqrt{3}/2)np(1 - 66np^2)$.

Recall $\Psi_1(u) \subset I_1$ for $u \in I_0$ and observe that,

$$\mathfrak{P}_u \leq \mathbb{E}[\mathbb{P}(\inf_{x \in I_1} d^*(x) < K_p \mid \mathfrak{F}_1)] + \mathbb{E}[\mathbb{P}(\varphi(x) \neq d^*(x) \text{ for some } x \in I_1 \mid \mathfrak{F}_1)],$$

by the tower property. Applying the union bound since $I_1 \in \mathfrak{F}_1$ yields

$$\mathfrak{P}_u \le \mathbb{E}[|I_1|\,\mathbb{P}(d^*(x) < K_p|x \in I_1, \mathfrak{F}_1)] + \mathbb{E}[|I_1|\,\mathbb{P}(\varphi(x) \neq d^*(x)|x \in I_1, \mathfrak{F}_1)].$$

Let $a := 13/\min\{c, 1\}$, where $c > 0$ is the largest real number such that $np \ge c \log n$. Separate the expectations into parts $\{|I_1| \le anp\}$ and $\{|I_1| > anp\}$ to give

$$\mathfrak{P}_u \le anp\mathbb{E}[\mathbb{P}(d^*(x) < K_p|x \in I_1, \mathfrak{F}_1)] + anp\mathbb{P}(\varphi(x) \neq d^*(x)|x \in I_1) + 2n\mathbb{P}(|I_1| > anp).$$

Since $d^*(x) \sim Bin(|S_1(x)|, p)$ by Lemma 3.1, $S_1(x) \in \mathfrak{F}_2$, and by Lemma 3.4 (i) we have

$$\mathfrak{P}_u \le anp\mathbb{E}[\mathbb{P}(Bin(|S_1(x)|, p) < K_p|\mathfrak{F}_2)] + a(np)e^{-(1-o(1))np} + 4n\mathbb{P}(d(u) > anp/2).$$

Applying Lemma A.1 to the first term and Lemma 3.2 with $\alpha = a$ to the last yields

$$\mathfrak{P}_u \le anp\mathbb{E}[e^{-(|S_1(x)|p-K_p)^2/(2|S_1(X)|p)}] + anp\cdot e^{-(1-o(1))np} + 4n\cdot o(e^{-anp/6}).$$

Once again by separating the expectation into the two disjoint parts $\{|S_1(x)| \le n-12(np)^2\}$ and $\{|S_1(x)| > n - 12(np)^2\}$ the applying Lemma 3.2 with $\alpha = 6$ we have the following:

$$\mathfrak{P}_u \le anp \cdot e^{-np/3} + 2\mathbb{P}(|B_2(u)| > 6(np)^2) + o(e^{-np}) = o(e^{-np/4}). \tag{44}$$

Recall $\sup\limits_{x\in\Psi_1(u)} \mathbf{1}_{\mathcal{B}_u^{u,v}}/\varphi(x) < 1/D$, see (20) and (22). Bernoulli's inequality (A1) provides

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}\left[\left(\sup_{x\in\Psi_1(u)}\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\varphi(x)}\right)^{\alpha}\right]^{1/\alpha} &\le \left(\frac{1}{(K_p)^\alpha} + \frac{1}{D^\alpha\mathbb{P}(\mathcal{C})}\mathbb{P}\left(\inf_{x\in\Psi_1(u)}\varphi(x) < K_p\right)\right)^{1/\alpha}\\ &\le \frac{1}{(K_p)}\left(1 + (K_p)^\alpha e^{-np/4}/D^\alpha\mathbb{P}(\mathcal{C})\right)^{1/\alpha} = O\left(\frac{1}{np}\right).\end{aligned} \tag{45}$$

Note that the bound (44) on $\mathfrak{P}_u$ holds for any $np \ge c \log n$, $c > 0$ fixed. The restriction on np to $np \ge \log n$ comes from (45), where we need $\mathbb{P}(\mathcal{C})$ bounded below by a constant.

Item (ii): Observe that $K_p = (1 - \sqrt{3}/2)np(1 - 66np^2) \ge np/9$ for large n and $k \le np\varepsilon_n$ in the definition of event $\mathcal{A}_{u,v}$ (24). Thus conditional on $\{\varphi(x) \ge K_p \text{ forall } x \in \Psi_1\} \cap \mathcal{A}_{u,v}$,

$$R(u, v) \le (1/\psi(u) + 1/\psi(v))(1 + (k + 2)/K_p) \le (1/\psi(u) + 1/\psi(v))(1 + 9\varepsilon_n),$$

by Corollary 2.5. The result follows since we have $\mathbb{P}(\inf_{x\in\Psi_1(u)}\varphi(x) < K_p) = o(e^{-np/4})$ by (44) and $\mathbb{P}((\mathcal{A}_{u,v})^c) = o(n^{-7/2})$ by Lemma 3.5. □

4.1 | Proof of Theorems 1.1 and 1.3

To begin let $i, j \in V$ and define the following three functions for ease of notation:

$$r_{i,j} := 1/d(i) + 1/d(j), \quad f_{i,j} := 1/d(i)^2 + 1/d(j)^2, \quad g_{i,j} := \varepsilon_n\cdot r_{i,j}.$$

Define the four events $\mathcal{I}, \mathcal{H}$ $\mathcal{F}$, and $\mathcal{L}$ as follows, where event $\mathcal{L}$ is w.r.t. MBFS($\mathcal{G}, \{i, j\}$),

$$\mathcal{I} := \bigcap_{\{i,j\}\subseteq V} \left\{ \left| R(i,j) - \frac{2}{np} \right| \le \frac{7\sqrt{\log n}}{(np)^{3/2}} \right\}, \quad \mathcal{H} := \{|R(i,j) - r_{i,j}| \le 9(2f_{i,j} + g_{i,j})\},$$
$$\mathcal{F} := \left\{ R(i,j) \le \left(\frac{1}{\psi(i)} + \frac{1}{\psi(j)} \right)(1 + 9\varepsilon_n) \right\}, \quad \mathcal{L} := \bigcap_{z\in\{i,j\}} \left\{ \frac{1}{d^*(z)} \le \frac{1}{d(z)} + \frac{16}{d(z)^2} \right\}. \tag{46}$$

Lemma 4.2. *Let $c \log n \le np \le n^{1/10}$, where $c > 0$. Then $\mathbb{P}(\mathcal{L}^c) = o(1/n^3) + o(e^{-np/3})$*

Proof. Let $I_0 = \{i, j\}$ and recall that $\Gamma^*(i) = \Gamma(i)\backslash(\{j\} \cap \Gamma(j))$, see (17) and (19). Let $\mathcal{E}_a = \{d^*(v) \le d(v) - a\}$. For each $z \in \{j\} \cap \Gamma(y)$, provided $z \ne i$, we have $iz \in E$ independently with probability p. Thus $\mathbb{P}(\mathcal{E}_a | ij \notin E, d(j) = k) = \mathbb{P}(Bin(k, p) \ge a)$ and similarly $\mathbb{P}(\mathcal{E}_a | ij \in E, d(j) = k) = \mathbb{P}(Bin(k - 1, p) \ge a - 1)$. Thus we have

$$\mathbb{P}(\mathcal{E}_a) \le \mathbb{P}(Bin(k, p) \ge a - 1) + \mathbb{P}(d(x) \ge k).$$

Let $k = 3a \max\{\log n, np\}$ and apply Lemma A.1 (ii) and (i), respectively to give

$$\mathbb{P}(\mathcal{E}_a) \le e^{-kp}\left(\frac{ekp}{a-1}\right)^{a-1} + e^{-\frac{k^2}{2(np+k/3)}} \le \left(\frac{2e(np)^2}{n}\right)^{a-1} + n^{-3a^2/2} = n^{-a/2}. \tag{47}$$

Conditional on the event $\mathcal{E}_j^c \cap \{d(v) \ge 2a\}$ we have

$$\frac{1}{d^*(x)} \le \frac{1}{d(x) - j} \le \frac{1}{d(x)} + \frac{j}{d(x)^2 - jd(x)} \le \frac{1}{d(x)} + \frac{2j}{d(x)^2}.$$

If we let $a = 8$ then $\mathbb{P}(d(v) < 16) \le o(e^{-np/3})$ by Lemma A.1 and $\mathbb{P}(\mathcal{E}_8) = o(1/n^3)$ by (47). □

Proof of Theorem 1.1. To begin, by Lemma 2.1 we have

$$R(i,j) - r_{i,j} \ge -\left(\frac{1}{d(i)^2 + d(i)} + \frac{1}{d(j)^2 + d(j)} \right) > -\left(\frac{1}{d(i)^2} + \frac{1}{d(j)^2} \right) = -f_{i,j}.$$

Let $\mathcal{B}$ be the event $\{\psi(i) = d^*(i), \psi(j) = d^*(j)\}$. Conditional on $\mathcal{B} \cap \mathcal{L} \cap \mathcal{F}$ we have

$$R(i,j) - r_{i,j} \le \frac{16}{d(i)^2} + \frac{16}{d(j)^2} + \left(\frac{1}{d(i)} + \frac{1}{d(j)} + \frac{16}{d(i)^2} + \frac{16}{d(j)^2} \right) 9\epsilon_n \le 9(2f_{i,j} + g_{i,j}).$$

Bounding $\mathbb{P}(\mathcal{B}^c)$, $\mathbb{P}(\mathcal{L}^c)$, and $\mathbb{P}(\mathcal{F}^c)$ using Lemmas 3.4 (ii), 4.2, and 5.1 (ii), respectively:

$$\mathbb{P}(\mathcal{H}^c) \le 2e^{-(1-o(1))np} + o(1/n^3) + o(e^{-np/3}) + o\ (e^{-np/4}) + o\ (n^{-7/2}),$$

which is $o\ (e^{-np/4}) + o\ (n^{-3})$ as required. □

For $S \subseteq V$ and $\lambda := \lambda(n) = o(np)$ let $\mathcal{E}(S, \lambda)$ be the event

$$\mathcal{E}(S, \lambda) := \bigcap_{u \in S} \{|d(u) - np| \le \sqrt{\lambda np}\}, \tag{48}$$

for which we have $\mathbb{P}(\mathcal{E}(S, \lambda)^c) \le 2|S|\, e^{-\lambda(n)/}$ by the union bound and Lemma A.1.

Proof of Theorem 1.3.
Item (i): Conditional on the event $\mathcal{E}(\{i, j\}, \varepsilon_n{}^2 np/19) \cap \mathcal{H}$ we have

$$\left| R(i,j) - \frac{2}{np} \right| \le |R(i,j) - r_{i,j}| + \left| r_{i,j} - \frac{2}{np} \right| \le \frac{19\varepsilon_n}{2np} + \frac{2\sqrt{\lambda(n)}}{(np)^{3/2}} \le \frac{10\varepsilon_n}{np}, \tag{49}$$

since $9(2f_{i,j} + g_{i,j}) \le 19\varepsilon_n/(2np)$ on $\mathcal{E}(\{i, j\}, \lambda) \cap \mathcal{H}$. Thus by Theorem 1.1

$$\mathbb{P}(|R(i,j) - 2/(np)| > 10\varepsilon_n/(np)) \le \mathbb{P}((\mathcal{H} \cap \mathcal{E})^c) \le \mathbb{P}(\mathcal{H}^c) + 4e^{-\lambda(n)/3} \le e^{-\varepsilon_n{}^2 np/60}.$$

Item (ii): Recall the definition of $\mathcal{H}$ from (46) and notice we suppressed dependence on i, j that is $\mathcal{H} := \mathcal{H}_{i,j}$. Similarly to (49), conditional on $(\cap_{\{i,j\} \subseteq V} \mathcal{H}_{i,j}) \cap \mathcal{E}(V, 9 \log n)$ we have

$$\left| R(i,j) - \frac{2}{np} \right| \le 2 \cdot \frac{3\sqrt{\log n}}{(np)^{3/2}} + \frac{19\varepsilon_n}{2np} \le \frac{7\sqrt{\log n}}{(np)^{3/2}}.$$

Recall event $\mathcal{I}$ from (46). The result now follows since by Theorem 1.1 and (48) we have

$$\mathbb{P}(\mathcal{I}^c) \le n^2(o(e^{-np/4}) + o(n^{-3})) + 2ne^{-3\log n} = o\left(\frac{1}{n}\right).$$

Item (iii): Recall that $m := |E|$ and let $\mathcal{M}$ be the event $\{|m - \binom{n}{2}p| \le 3\sqrt{\log(n)\binom{n}{2}p}\}$. Conditional on $\mathcal{E}(V, 9 \log n) \cap \mathcal{M} \cap \mathcal{I}$ we have the following for any $\{i, j\} \subseteq V$:

$$|mR(i,j) - n| \le 4n\sqrt{\frac{\log n}{np}}, \quad \text{and} \quad \left| \sum_{u \in V} \frac{d(u)}{2} [R(j,u) - R(u,i)] \right| \le 8n\sqrt{\frac{\log n}{np}},$$

thus $|h(i,j) - n| \le 12\sqrt{\log(n)/np}$ by Tetali's formula (3) and the Triangle inequality. Now

$$\mathbb{P}((\mathcal{E}(V, 9 \log n) \cap \mathcal{M} \cap \mathcal{I})^c) = o(1/n^3) + o(1/n^3) + o(1/n) = o(1/n)$$

by (48), Lemma A.1, since $m \sim Bin(\binom{n}{2}, p)$ and Theorem 1.3 (ii) respectively. □

4.2 | Proof of Theorem 1.6

Recall that $paths_2(i, j, l)$ is the maximum number of paths of length at most l between vertices i and j that are vertex disjoint on $V \backslash (B_1(i) \cup B_1(j))$ of a graph G.

Proof of Theorem 1.6.

Item (i): For $i, j \in V$ let $\mathcal{E}_{i,j}$ be the event that there is no path from i to j of length less than 4. Then by over-counting the number of paths we have

$$\mathbb{P}\left(\mathcal{E}_{i,j}^c\right) \leq \sum_{l=1}^{3} \mathbb{P}(\text{path from } i \text{ to } j \text{ of length } l) \leq p + (n-2)p^2 + \binom{n-2}{2} p^3 \leq n^2 p^3. \tag{50}$$

Conditional on $\mathcal{E}_{i,j}$ every path between i and j must pass through at least one vertex from each of $d_2(i)$ and $d_2(j)$, though these vertices may not be distinct. So there cannot be more than $\min\{d_2(i), d_2(j)\}$ paths between $i, j \in V$ which are vertex disjoint on $V^* := V \backslash (B_1(i) \cup B_1(j))$ since $\Gamma_2(i) \cup \Gamma_2(j) \subseteq V^*$. Thus conditional on $\mathcal{E}_{i,j}$ for any $l \geq 0$ we have

$$paths_2(i, j, l) \leq \min\{d_2(i), d_2(j)\}. \tag{51}$$

To bound $paths_2(i, j, l)$ from below we construct $\min\{\psi_2(i), \psi_2(j)\}$ vertex disjoint paths between i and j conditional on $\mathcal{A}_{i,j}$, Definition 2.4, then couple $\psi_2(i)$ to $d_2(i)$ and $\psi_2(j)$ to $d_2(j)$.

For the path construction condition on $\mathcal{A}_{i,j}$ and w.l.o.g. assume $\psi_2(i) \leq \psi_2(j)$. Take any subset $\Psi_2(j)^* \subseteq \Psi_2(j)$ with $\psi_2(i)$ elements and any bijection M between $\Psi_2(i)$ and $\Psi_2(j)^*$. Given any pair (x, y) in M, conditional on $\mathcal{A}_{i,j}$, there is some k and some pair $(x_k, y_k) \in \Gamma_k^*(x) \times \Gamma_k^*(y)$ such that $x_k y_k \in E$. We define the path $P_{x,y} := i, i_x, x, x_1, \dots, x_k, y_k, y_{k-1}, \dots, y, j_y, j$, where $x, x_1, \dots, x_k$ is the unique path from x to x_k in the tree $T_k(x) := \cup_{i=0}^k \Gamma_i^*(x)$ and i_x is the unique vertex in $\Gamma_1^*(i)$ connected to x. The equivalent descriptions hold for $y, y_1, \dots, y_k \in T_k(y)$ and $j_y \in \Gamma_1(j)$ with respect to y and j. The paths $\{P_{x,y}\}_{(x,y)\in M}$ are all vertex disjoint on V^* since the trees $\{T_K(u)\}_{u \in \Psi_2}$ are all vertex disjoint. Each path in $P_{i,j}$ has length $l := 2k + 5$, where the k is given by the event $\mathcal{A}_{i,j}$. Thus conditional on the event $\mathcal{A}_{i,j}$ we have

$$paths_2(i, j, l) \geq |\{P_{x,y}\}_{(x,y)\in M}| = \min\{\psi_2(i), \psi_2(j)\}. \tag{52}$$

Exchanging the ψ_2 and d_2 distributions on the event $\{\psi_2(i) \neq d_2(i) \text{ or } \psi_2(j) \neq d_2(j)\}$ yields

$$\begin{aligned}\mathfrak{P} := \mathbb{P}(paths_2(i, j, l) \neq \min\{d_2(i), d_2(j)\}) &\leq \mathbb{P}(\psi_2(i) \neq d_2(i) \text{ or } \psi_2(j) \neq d_2(j)) \\ &+ \mathbb{P}(paths_2(i, j, l) < \min\{\psi_2(i), \psi_2(j)\}) + \mathbb{P}(\{paths_2(i, j, l) > \min\{d_2(i), d_2(j)\}\}).\end{aligned}$$

Now by (52) and (51) we have the following:

$$\begin{aligned}\mathfrak{P} \;&\leq\; \mathbb{P}(\psi_2(i) \neq d_2(i) \text{ or } \psi_2(j) \neq d_2(j)) + \mathbb{P}((\mathcal{A}_{i,j})^c) + \mathbb{P}(\mathcal{E}_{i,j}^c) \\ &\leq\; 5n^3p^4 + o(e^{-7 \min\{np, \log n\}/2}),\end{aligned}$$

by Lemma 3.4 (v), Lemma 3.5 and (50), respectively. On the event $\mathcal{A}_{i,j}$ the strong path property is satisfied for some $k \leq \lfloor \frac{\log n}{2\log(np)} \rfloor + 2$, thus $l = 2k + 5 \leq \frac{\log n}{\log(np)} + 9$.

Item (ii): Observe that $d_2(u) \sim Bin(n - 1 - d(u), 1 - (1-p)^{d(u)})$, conditional on $d(u)$ for any $u \in V$. Notice that $(1-p)^k \leq 1 - kp + (kp)^2$ when $(kp)^i \geq (kp)^{i+1}$ for all i by the Bernoulli inequality (A1). Thus conditional on $\mathcal{E}(\{i, j\}, 3\log(np))$, see (48), we have the following:

$$Bin\Big(n - 2np, np^2 - 2p\sqrt{\log(np)np}\Big) \preccurlyeq_1 d_2(i), d_2(j) \preccurlyeq_1 Bin\Big(n, np^2 + p\sqrt{3\log(np)np}\Big).$$

Let $\mathcal{R}_{i,j}$ be the event $\{|\min\{d_2(i), d_2(j)\} - (np)^2| \le 3(np)^{3/2}\sqrt{\log np}\}$. Observe that we have

$$\mathbb{P}\Big(\mathcal{R}_{i,j}^c\Big) \le \mathbb{P}\Big(\mathcal{R}_{i,j}^c \Big| \mathcal{E}(\{i,j\}, 3\log(np))\Big) + \mathbb{P}(\mathcal{E}(\{i,j\}, 3\log(np))^c) = o(1/np), \tag{53}$$

by (48) and applying Chernoff bounds to $d_2(i)$ conditional on $\mathcal{E}(\{i,j\}, 3\log(np))$. We now have

$$\begin{aligned}\mathbb{P}(|paths_2(i,j,l) - (np)^2| > 3(np)^{3/2}\sqrt{\log np}) &\le \mathbb{P}(paths_2(i,j,l) \ne \min\{d_2(i), d_2(j)\}) \\ + \mathbb{P}\Big(\mathcal{R}_{i,j}^c\Big) &\le 5n^3p^4 + o(e^{-7\min\{np, \log n\}/2}) + o(1/np) = o(1/np),\end{aligned}$$

by Item (i) and the bound on $\mathbb{P}(\mathcal{R}_{i,j}^c)$ from (53). □

5 | PROOF OF THEOREMS 1.2, 1.4, 1.7, AND 1.9

Recall $\varepsilon_n := \frac{\log n}{np\log(np)}$ from (2), that $m = |E|$ and Tetali's formula (3), which is given by

$$h(i,j) = mR(i,j) + \sum_{u \in V} \frac{d(u)}{2}[R(j,u) - R(u,i)].$$

Our results on hitting times and other random walk indices come from applying our bounds on resistance to Tetali's formula (3) to obtain moments hitting times. The following two Lemmas help us calculate the terms arising during these computations.

Lemma 5.1. *Let $\mathcal{G} \sim \mathcal{G}(n,p)$, where $\log n + \log\log\log n \le np < o(n^{1/3})$. Let $\alpha \ge 1$ and $\Psi_1(u), \Psi_1(v)$ be defined with respect to* MBFS$(\mathcal{G}, \{u,v\})$, *$u, v \in V$. Then*

$$\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\psi(u)^\alpha}\right]^{1/\alpha} = \frac{1 + O(\epsilon_n)}{np}.$$

Proof. We restrict to the event $\mathcal{B}_u^{u,v}$ to ensure the expectation is bounded,

$$\mathfrak{E} := \mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{B}_u^{u,v}}}{\psi(u)^\alpha}\right] = \sum_{k=1}^{n} \frac{1}{k^\alpha}\mathbb{P}\mathcal{C}(\psi(u) = k) = \sum_{k=1}^{n} \frac{1}{k^\alpha}\frac{\mathbb{P}(\{\psi(u) = k\} \cap \mathcal{C})}{\mathbb{P}(\mathcal{C})}.$$

Applying the coupling inequality (A3), and then Lemma 3.4 to bound $\mathbb{P}(d^*(u) \ne \psi(u))$ gives

$$\mathfrak{E} \le \sum_{k=1}^{n} \frac{1}{k^\alpha}\frac{\mathbb{P}(d^*(u) = k) + \mathbb{P}(d^*(u) \ne \psi(u))}{\mathbb{P}(\mathcal{C})} = \sum_{k=1}^{n} \frac{1}{k^\alpha}\frac{\mathbb{P}(d^*(u) = k) + e^{-(1-o(1))np}}{\mathbb{P}(\mathcal{C})}.$$

Let $\tilde{d}_1(v) := |\Gamma_1(v) \cap S_0| \sim Bin(n-2, p)$. By Lemma 3.1 we have $d^*(u) \sim Bin(n-2-h, p)$ conditional on $\{\tilde{d}_1(v) = h\}$. By the law of total expectation and the generalised harmonic series,

$$\mathfrak{E} \leq \sum_{k=1}^{n} \frac{1}{k^\alpha} \frac{\sum_{h=0}^{n-2} \mathbb{P}(d^*(u) = k \,|\, \tilde{d}_1(v) = h)\mathbb{P}(\tilde{d}_1(v) = h)}{\mathbb{P}(\mathcal{C})} + O\left(\frac{(\log n) e^{-(1-o(1))np}}{\mathbb{P}(\mathcal{C})}\right).$$

Now by writing out $\mathbb{P}(d^*(u) = k \,|\, \tilde{d}_1(v) = h)\mathbb{P}(\tilde{d}_1(v) = h)$ explicitly we have

$$\begin{aligned}
\mathfrak{E} &\leq \sum_{k=1}^{n} \frac{1}{k^\alpha} \frac{\sum_{h=0}^{n-2} \binom{n-2-h}{k} p^k (1-p)^{n-2-h-k} \cdot \binom{n-2}{h} p^h (1-p)^{n-2-h}}{\mathbb{P}(\mathcal{C})} + e^{-(1-o(1))np} \\
&= \sum_{h=0}^{n-3} \binom{n-2}{h} \frac{p^h (1-p)^{n-2-h}}{\mathbb{P}(\mathcal{C})} \left(\sum_{k=1}^{n-2-h} \frac{1}{k^\alpha} \binom{n-2-h}{k} p^k (1-p)^{n-2-h-k} \right) + e^{-(1-o(1))np}.
\end{aligned}$$

Applying Proposition A.3 to the bracketed sum above where we let B_h be a random variable with distribution $Bin(n-h-3, p)$ yields

$$\mathfrak{E} \leq \frac{np}{\mathbb{P}(\mathcal{C})} \sum_{h=0}^{n-3} \binom{n-2}{h} p^h (1-p)^{n-2-h} \mathbb{E}\left[\frac{1}{(B_h+1)^{\alpha+1}}\right] + e^{-(1-o(1))np}.$$

The weight in front of the expectation term is the density of a $Bin(n-2, p)$ random variable. Split the sum at $t := \sqrt{3np(\alpha+2)\log(np)}$ and bound the expectation to give

$$\mathfrak{E} \leq \frac{np}{\mathbb{P}(\mathcal{C})} \left(\mathbb{P}(Bin(n-2, p) \leq t) \mathbb{E}\left[\frac{1}{(B_t+1)^{\alpha+1}}\right] + \mathbb{P}(Bin(n-2, p) > t) \right) + e^{-(1-o(1))np}.$$

Bounding $\mathbb{P}(Bin(n-2, p) > t)$ by Lemma A.1 using Lemma A.3 to calculate $\mathbb{E}\left[\frac{1}{(B_t+1)^{\alpha+1}}\right]$:

$$\begin{aligned}
\mathfrak{E} &\leq \frac{np}{\mathbb{P}(\mathcal{C})} \left[\left(\frac{1}{((n-t-3)p)^{\alpha+1}} + O\left(\frac{1}{((n-t-3)p)^{\alpha+2}}\right) \right) + O\left(\frac{1}{(np)^{\alpha+2}}\right) \right] + e^{-(1-o(1))np} \\
&= \frac{1}{\mathbb{P}(\mathcal{C})} \left(\frac{1}{(np)^\alpha} + O\left(\frac{1}{(np)^{\alpha+1}}\right) \right)
\end{aligned}$$

Applying Bernoulli's inequality (A1) yields

$$\mathfrak{E}^{1/\alpha} \leq \frac{1}{\mathbb{P}(\mathcal{C})} \left(\frac{1}{(np)^\alpha} + O\left(\frac{1}{(np)^{\alpha+1}}\right) \right)^{1/\alpha} = \frac{1}{1 - \mathbb{P}(\mathcal{C}^c) np} \left(1 + O\left(\frac{1}{np}\right) \right)^{1/\alpha} = \frac{1 + O(\epsilon_n)}{np},$$

as (10) gives $\mathbb{P}(\mathcal{C}^c) \leq O(\varepsilon_n)$ whenever $np \geq \log n + \log\log\log n$. □

Lemma 5.2. *For any set $A \subset V$ of size $0 \le a \le 3$ and any set of vertex pairs $B \subset \binom{V}{2}$ of size $0 \le b \le 3$ then*

$$\mathbb{E}_{\mathcal{C}}\left[\left(\prod_{v\in A} d(v)\right)\left(\prod_{\{x,y\}\in B} R(x,y)\right)\right] = \frac{2^b}{(np)^{a-b}}(1 \pm O(\varepsilon_n)).$$

Proof. We shall prove the case $A = \{u, v, w\}$ and $B = \{(a_1, a_2), (b_1, b_2), (c_1, c_1)\}$, this is the "largest" case and the other cases are proved in exactly the same way. Let

$$\mathcal{E} := \mathcal{A}^n_{a_1,a_2} \cap \mathcal{A}^n_{b_1,b_2} \cap \mathcal{A}^n_{c_1,c_2} \cap \mathcal{B}_{a_1,a_2} \cap \mathcal{B}_{b_1,b_2} \cap \mathcal{B}_{c_1,c_2}.$$

For ease of notation we define

$$\mathsf{Deg}(A) := \prod_{v\in A} d(v) \quad \text{and} \quad \mathsf{Res}(B); = \prod_{\{x,y\}\in B} R(x,y).$$

Recall the bound on $R(x, y)$ from Corollary 2.5, conditional on $\mathcal{A}_{x,y}$, this yields

$$\begin{aligned}\mathrm{Res}(B)\mathbf{1}_{\mathcal{E}} &\le \prod_{\{x,y\}\in B}\left(\frac{1}{\psi(x)} + \frac{1}{\psi(y)} + \frac{k+2}{\psi(x)}\sup_{a\in\Psi_1(x)}\frac{1}{\varphi(a)} + \frac{k+2}{\psi(y)}\sup_{b\in\Psi_1(y)}\frac{1}{\varphi(b)}\right)\mathbf{1}_{\mathcal{E}} \\ &= \sum_{\substack{x\in\{a_1,a_2\}\\ y\in\{b_1,b_2\}\\ z\in\{c_1,c_2\}}}\left(a_{x,y,z} + \sum_{\substack{f,g,h\in\\ \{x,y,z\}\\ f\neq g\neq h}}[(k+2)\cdot b_{f,g,h} + (k+2)^2\cdot c_{f,g,h}] + (k+2)^3\cdot d_{x,y,z}\right)\end{aligned} \tag{54}$$

where the summands are given by

$$a_{x,y,z} = \frac{\mathbf{1}_{\mathcal{E}}}{\psi(x)\psi(y)\psi(z)}, \qquad b_{f,g,h} = \frac{\mathbf{1}_{\mathcal{E}}}{\psi(f)\psi(g)\psi(h)}\sup_{a\in\Psi_1(h)}\frac{1}{\varphi(a)},$$
$$c_{f,g,h} = \frac{\mathbf{1}_{\mathcal{E}}}{\psi(f)}\prod_{w\in\{g,h\}}\frac{1}{\psi(w)}\sup_{a\in\Psi_1(w)}\frac{1}{\varphi(a)}, \qquad d_{x,y,z} = \mathbf{1}_{\mathcal{E}}\prod_{w\in\{f,g,h\}}\frac{1}{\psi(w)}\sup_{a\in\Psi_1(w)}\frac{1}{\varphi(a)}.$$

By Hölder's inequality (A2), it follows that $\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot a_{x,y,z}\cdot \mathbf{1}_{\mathcal{E}}]$ is at most

$$\begin{aligned}&\mathbb{E}_{\mathcal{C}}[d(u)^6]^{\frac16}\,\mathbb{E}_{\mathcal{C}}[d(v)^6]^{\frac16}\,\mathbb{E}_{\mathcal{C}}[d(w)^6]^{\frac16}\,\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{E}}}{\psi(x)^6}\right]^{\frac16}\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{E}}}{\psi(y)^6}\right]^{\frac16}\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{E}}}{\psi(z)^6}\right]^{\frac16} \\ &= ((np)^6 + O((np)^5))^{\frac12}\cdot\left(\frac{1+O(\varepsilon_n)}{np}\right)^3 = 1 + O(\varepsilon_n),\end{aligned}$$

where we applied (A4) and Lemma 5.1 to the expectations, then Bernoulli's inequality (A1).

Similarly by Hölder's inequality (A2) and collecting similar terms

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot b_{f,g,h}\cdot \mathbf{1}_{\mathcal{E}}] \le \mathbb{E}_{\mathcal{C}}[d(u)^7]^{\frac{3}{7}}\,\mathbb{E}_{\mathcal{C}}\left[\frac{\mathbf{1}_{\mathcal{C}_1}}{\psi(f)^7}\right]^{\frac{3}{7}}\mathbb{E}_{\mathcal{C}}\left[\sup_{c\in\Psi_1(h)}\frac{\mathbf{1}_{\mathcal{C}_1}}{\varphi(c)^7}\right]^{\frac{1}{7}}$$
$$= ((np)^7 + O((np)^6))^{\frac{3}{7}}\cdot\left(\frac{1+O(\varepsilon_n)}{np}\right)^3\cdot O\left(\frac{1}{np}\right) = O\left(\frac{1}{np}\right).$$

where in addition we applied Lemma 4.1. By a near identical calculation we have

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot c_{f,g,h}\cdot \mathbf{1}_{\mathcal{E}}] = O\left(\frac{1}{(np)^2}\right),$$
$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot d_{x,y,z}\cdot \mathbf{1}_{\mathcal{E}}] = O\left(\frac{1}{(np)^3}\right).$$

Now by linearity of expectation, (54), and since $k = O(\log(n)/\log(np))$, we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot \mathsf{Res}(B)\cdot \mathbf{1}_{\mathcal{E}}] &= 2^3 + O(\varepsilon_n) + O\left(\frac{k}{np}\right) + O\left(\frac{k^2}{(np)^2}\right) + O\left(\frac{k^3}{(np)^3}\right)\\ &= 2^3 + O(\varepsilon_n).\end{aligned} \tag{55}$$

We shall now consider what happens on $\mathcal{E}^c$, let $\mathcal{M}$ be the event $\cap_{u\in A}\{d(u)\le 8np\}$. By Chernoff bounds Lemma A.1 and the bound (10) on $\mathbb{P}(\mathcal{C})$ we have $\mathbb{P}_{\mathcal{C}}(\mathcal{M}^c) = o(1/n^7)$. Let $\mathcal{S}_{i,j}$ be the event $\{R(i,j)\le 3\log n/\log(np)\}$ and recall $\mathbb{P}_{\mathcal{C}}(\mathcal{S}_{i,j}^c) = o(n^{-4})$ by Lemma 3.6.

Observe that conditional on $\tilde{\mathcal{E}}_1 = \mathcal{E}^c\cap\mathcal{M}\cap\prod_{\{x,y\}\in B}\mathcal{S}_{x,y}$ the following inequalities hold for all $v\in A$ and $\{x,y\}\in B$: $d(u)\le 8np$ and $R(x,y)\le 3\log(n)/\log(np)$. Thus

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot \mathsf{Res}(B)\cdot \mathbf{1}_{\tilde{\mathcal{E}}_1}] = O\left((np)^3\cdot\frac{(\log n)^3}{\log(np)}\right)\cdot\mathbb{P}_{\mathcal{C}}(\mathcal{E}^c) = o(1/n^{4/5}). \tag{56}$$

We shall now consider conditioning on the event $\tilde{\mathcal{E}}_2 = \mathcal{E}^c\cap\mathcal{M}\cap(\prod_{\{x,y\}\in B}\mathcal{S}_{x,y})^c$ where we instead use the worse case resistance bound $R(i,j)\le n-1$, this gives

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot \mathsf{Res}(B)\cdot \mathbf{1}_{\tilde{\mathcal{E}}_2}] = O((np)^3\cdot n^3)\cdot\mathbb{P}_{\mathcal{C}}(\mathcal{S}_{x,y}^c) = o(1/n^{4/5}). \tag{57}$$

Finally we consider the event $\mathcal{E}^c\cap\mathcal{M}^c$ and we observe that since $\mathbb{P}_{\mathcal{C}}(\mathcal{M}^c) = o(1/n^7)$ we have

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot \mathsf{Res}(B)\cdot \mathbf{1}_{\mathcal{E}^c\cap\mathcal{M}^c}] = O(n^6)\cdot\mathbb{P}_{\mathcal{C}}(\mathcal{M}^c) = o(1/n). \tag{58}$$

The upper bound on $\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot\mathsf{Res}(B)]$ follows by combining (55), (56), (57), and (58).

We now consider the lower bound. Recall that $B = \{(a_1,a_2),(b_1,b_2)(c_1,c_2)\}$, in the case we are considering. Lemma 2.1 states $R(x,y)\ge 1/(d(x)+1)+1/(d(y)+1)$, thus

$$\mathbb{E}_{\mathcal{C}}[\mathsf{Deg}(A)\cdot\mathsf{Res}(B)] \geq \sum_{x,y,z} \mathbb{E}_{\mathcal{C}}\left[\frac{\prod_{u\in A} d(u)}{(d(x)+1)(d(y)+1)(d(z)+1)}\right], \tag{59}$$

where the sum is over $(x, y, z) \in \{a_1, a_2\} \times \{b_1, b_2\} \times \{c_1, c_2\}$. Let $\mathcal{D}$ be the event given by

$$\mathcal{D} := \left(\prod_{u\in A}\{d(u) \geq np - a\sqrt{np}\}\right) \bigcap \left(\prod_{\{x,y\}\in B}\{d(x), d(y) \leq np + a\sqrt{np}\}\right),$$

where $a = 3\sqrt{\log\ \log\ n}$ if $np = O(\log n)$ and $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$. Then,

$$\mathbb{E}_{\mathcal{C}}\left[\frac{\prod_{u\in A} d(u)}{(d(x)+1)(d(y)+1)(d(z)+1)}\right] \geq \frac{(np - a\sqrt{np})^3}{(np + a\sqrt{np})^3}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) = 1 - O\ (\varepsilon_n),$$

where the bound on $\mathbb{P}_{\mathcal{C}}(\mathcal{D})$ is by Lemma A.1. The lower bound follows from (59). □

5.1 | Proof of Theorem 1.2

Equipped with Lemma 5.2, the proofs of the main "moment theorems" are straightforward.

Proof of Theorem 1.2. Observe that $\mathbb{E}_{\mathcal{C}}[R(i, j)] = (2 \pm O\ (\varepsilon_n))/np$ follows directly from Lemma 5.2 with $A = \varnothing$ and $B = \{(i, j)\}$. For hitting times we have the following by (3):

$$\mathbb{E}_{\mathcal{C}}[h(i,j)] = \mathbb{E}\mathcal{C}[mR(i,j)] + \frac{1}{2}\sum_{u\in V}(\mathbb{E}_{\mathcal{C}}\ [d(u)R(u,j)] - \mathbb{E}_{\mathcal{C}}\ [d(u)R(u,i)]) = \mathbb{E}_{\mathcal{C}}[mR(i,j)],$$

when $i \neq j$, by symmetry. Thus, we have

$$\mathbb{E}_{\mathcal{C}}[h(i,j)] = \frac{1}{2}\sum_{u\in V}\mathbb{E}_{\mathcal{C}}[d(u)R(i,j)] = \frac{1}{2}\sum_{u\in V}(2 \pm O(\varepsilon_n)) = n(1 \pm O\ (\varepsilon_n)),$$

by Lemma 5.2 with $A = \{u\}$ and $B = \{(i, j)\}$. □

5.2 | Proof of Theorem 1.7

Theorems 1.4 and 1.9 shall be proved by Chebychev's inequality, thus we need second moments.

Lemma 5.3. *Let $\mathcal{G} \sim \mathcal{G}(,n, p)$ satisfy* (4) *and $i, j \in V(\mathcal{G})$, where $i \neq j$. Then $\mathbb{E}_{\mathcal{C}}[h(i,j)^2] = (1 \pm O\ (\varepsilon_n))n^2$, $\mathbb{E}_{\mathcal{C}}[cc_i(\mathcal{G})^2] = (1 \pm O\ (\varepsilon_n))n^2$ and $\mathbb{E}_{\mathcal{C}}[K(\mathcal{G})^2] = (1 \pm O\ (\varepsilon_n))n^2/p^2$.*

Proof. Let $g(a, b, c, d) := \mathbb{E}_{\mathcal{C}}[d(u)d(v)R(a, b)R(c, d)]$. Using Tetali's formula (3) we can expand $\mathbb{E}_{\mathcal{C}}[h(i, j)h(i, a)]$ to give the following for any $i, j, a \in V$:

$$\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{u\in V}\frac{d(u)}{2}(R(i,j)+R(j,u)-R(u,i))\right)\cdot\left(\sum_{v\in V}\frac{d(v)}{2}(R(i,a)+R(a,v)-R(v,i))\right)\right]$$
$$=\frac{1}{4}\sum_{u,v\in V}\left(g(i,j,i,a)+\sum_{\substack{(w,z)\in\\ \{(u,i),(j,a)\}}}g(i,w,v,z)-\sum_{w\in\{i,u\}}g(w,j,i,v)\right)$$
$$+\frac{1}{4}\sum_{u,v\in V}\sum_{w\in\{i,v\}}(g(u,j,w,a)-g(w,a,i,u))=\frac{1}{4}\sum_{u,v\in V}\mathbb{E}_{\mathcal{C}}[d(u)d(v)R(i,j)R(i,a)]. \tag{60}$$

To see the above, observe that $R(a,b)R(c,d)=0$ if and only $a=b$ or $c=d$. Thus only the first term, $g(i,j,i,a)$, will always be non-zero. All the other terms contain one or more input from $\{u,v\}$ so will be zero at different times. Of the eight other terms there are two positive and two negative terms containing one of $\{u,v\}$, then two positive and two negative terms containing both u and v as inputs. Thus by symmetry when the sums are expanded everything apart from the first term $g(i,j,i,a)$ cancels.

Thus by (60) and Lemma 5.2 with $A=\{u,v\}$ and $B=\{(i,j),(i,a)\}$ we have

$$\mathbb{E}_{\mathcal{C}}[h(i,j)h(i,a)]=\frac{1}{4}\sum_{u,v\in V}(4\pm O(\epsilon_n))=n^2(1\pm O(\epsilon_n)). \tag{61}$$

Now by the definition (6) of $cc_i(G)$ and (61) we have,

$$\mathbb{E}_{\mathcal{C}}[cc_i(\mathcal{G})^2]=\frac{1}{(n-1)^2}\mathbb{E}_{\mathcal{C}}\left[\left(\sum_{j\in V}h(i,j)\right)^2\right]=\frac{1}{(n-1)^2}\sum_{j,k\in V;j,k\neq i}\mathbb{E}_{\mathcal{C}}[h(i,j)h(i,k)],$$

which is equal to $(1\pm O(\epsilon_n))n^2$. Finally observe that by (6) we have

$$\mathbb{E}_{\mathcal{C}}[K(\mathcal{G})^2]=\sum_{\{i,j\}\subseteq V}\sum_{\{w,z\}\subseteq V}\mathbb{E}_{\mathcal{C}}[R(i,j)R(w,z)]=\sum_{\{i,j\}\subseteq V}\sum_{\{w,z\}\subseteq V}\frac{4\pm O(\epsilon_n)}{(np)^2}=\frac{n^2}{p^2}(1\pm O(\epsilon_n)),$$

where we applied Lemma 5.2 with $A=\emptyset$ and $B=\{(i,j),(w,z)\}$. □

Proof of Theorem 5.2. Recall (5), the definitions of $H_i(G)$ for $i\in V$ and $T(G)$:

$$H_i(G):=\sum_{j\in V}\frac{d(j)}{2m}h(j,i),\qquad T(G):=\sum_{j\in V}\frac{d(j)}{2m}h(i,j),$$

where $m:=|E|\sim Bin(\binom{n}{2},p)$. To begin, let $m^*\sim Bin(\binom{n}{2}-1,p)$, $k\in\mathbb{Z}$, $k\geq 1$. Proposition A.3 and the fact that $\mathcal{C}\subset\{m\geq 1\}$ yields the following:

$$\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^k}\right]\mathbb{P}(\mathcal{C})=\mathbb{E}\left[\frac{\mathbf{1}_{\mathcal{C}}}{m^k}\right]\leq\mathbb{E}\left[\frac{\mathbf{1}_{\{m\geq 1\}}}{m^k}\right]=\mathbb{E}\left[\frac{\binom{n}{2}p}{(m^*+1)^{k+1}}\right]=\frac{2^k}{n^{2k}p^k}\left(1+O\left(\frac{1}{np}\right)\right),$$

where in the last step we used Lemma A.3 to bound the expectation term. Observe that by (10), $\mathbb{P}(\mathcal{C}^c) \le O(\varepsilon_n)$ whenever $np \ge \log n + \log\log\log n$. Thus by the Bernoulli inequality (A1) for any given $a, k \in \mathbb{Z}, a, k \ge 1$ we have

$$\mathbb{E}_{\mathcal{C}}\left[\frac{1}{m^k}\right]^{1/a} = \frac{2^{k/a}}{n^{2k/a}p^{k/a}}\left(1 + \frac{\mathbb{P}(\mathcal{C}^c)}{\mathbb{P}(\mathcal{C})} + O\left(\frac{1}{np}\right)\right)^{1/a} \le \frac{2^{k/a}}{n^{2k/a}p^{k/a}}(1 + O(\varepsilon n)). \tag{62}$$

Using Hölder's inequality to break the product of random variables in the expectation:

$$\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] \le (1/2)\sum_{j\in V}\mathbb{E}_{\mathcal{C}}[d(j)^4]^{1/4}\mathbb{E}_{\mathcal{C}}[1/m^4]^{1/4}\mathbb{E}_{\mathcal{C}}[h(i,j)^2]^{1/2}.$$

Then applying (A4), (62) and the upper bound on $\mathbb{E}_{\mathcal{C}}[h(i,j)^2]$ from Lemma 5.3 yields

$$\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] \le (n/2)((np)^4 + O((np)^3))^{1/4}\cdot[(2 + O(\varepsilon_n))/n^2p]\cdot n(1 + O(\varepsilon n)) = n(1 + O(\varepsilon n)).$$

The same upper bounds for $\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})]$ follows by identical steps. By (3) we have

$$T(G) = \sum_{j\in V}\frac{d(j)}{2m}\left(mR(i,j) + \sum_{u\in V}\frac{d(u)}{2}[R(u,j) - R(u,i)]\right)$$

for G connected. Applying the effective resistance bound, Lemma 2.1, and reducing yields

$$\begin{aligned} T(G) \ge{}& \frac{m}{(d(i)+1)} - \frac{d(i)}{2(d(i)+1)} + \sum_{\substack{j\in V\\ j\ne i}}\frac{d(j)}{2(d(j)+1)} + \sum_{\substack{j,u\in V\\ j\ne u}}\frac{d(j)d(u)}{2m(d(u)+1)} \\ &- \sum_{j,u\in V}\frac{d(u)d(j)}{4m}R(u,i). \end{aligned}$$

Applying $d(i)/(d(i)+1) = 1 - 1/(d(i)+1)$ and the bound $d(i)/(d(i)+1) \le 1$ yields

$$T(G) \ge \frac{m}{d(i)+1} + \frac{3n}{2} - 2 - \sum_{u\in V}\frac{3}{2(d(u)+1)} - \sum_{u\in V}\frac{d(u)}{2}R(u,i).$$

Again by a similar procedure we have the following for the stationary hitting time $H_i(G)$

$$\begin{aligned} H_i(G) ={}& \sum_{j\in V}\frac{d(j)}{2m}\left(mR(j,i) + \sum_{u\in V}\frac{d(u)}{2}[R(u,i) - R(u,j)]\right) \ge \frac{n-1}{2} - \sum_{j\in V}\frac{1}{2(d(j)+1)} \\ &+ \frac{m-1}{(d(i)+1)} - 1 + \sum_{u\in V, u\ne i}\frac{d(u)}{2}\left(\frac{1}{d(i)+1} + \frac{1}{d(u)+1}\right) - \sum_{j,u\in V}\frac{d(u)d(j)}{4m}R(u,j) \\ &\ge n + \frac{2m-2}{d(i)+1} - \sum_{u\in V}\frac{1}{d(u)+1} - \frac{7}{2} - \sum_{j,u\in V}\frac{d(u)d(j)}{4m}R(u,j). \end{aligned}$$

Let $\mathcal{D}$ be the event $\{m \geq n^2p/2 - a\sqrt{n^2p/2}\} \cap \{d(j) \leq np + a\sqrt{np}\}$, where $a = 3\sqrt{\log\ \log\ n}$ if $np = O(\log n)$ and $a = 3\sqrt{\log n}$ if $np = \omega(\log n)$. Now by Lemma A.1 we obtain

$$\mathbb{P}_{\mathcal{C}}(\mathcal{D}) = (1 - \exp(-a^2/2)/\mathbb{P}(\mathcal{C}) - \exp(-a^2/2(1 + a/3\sqrt{np}))/\mathbb{P}(\mathcal{C})) = 1 - o(1/np).$$

By Hölder's inequality (A2), $1 \geq \mathbf{1}_{\mathcal{D}}$ and the bound on $\mathbb{P}_{\mathcal{C}}(\mathcal{D})$ in the line above we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})] \geq n &+ 2\frac{\binom{n}{2}p - a\sqrt{\binom{n}{2}p} - 1}{np + a\sqrt{np} + 1}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) - n \cdot \mathbb{E}_{\mathcal{C}}\left[\frac{1}{d(u) + 1}\right] - \frac{7}{2}\\ &-(n/4)\mathbb{E}_{\mathcal{C}}[d(j)^4]^{1/4}\mathbb{E}_{\mathcal{C}}[1/m^4]^{1/4}\mathbb{E}_{\mathcal{C}}[d(u)^2R(u,j)^2]^{1/2} = n(1 - O(\varepsilon n)).\end{aligned}$$

The last equality comes from applying estimates to the expectation terms which are given by Lemma A.3 in Appendix A and (A4), (62), and Lemma 5.2, respectively. Similarly we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})] \geq &\frac{\binom{n}{2}p - 2a\sqrt{\binom{n}{2}p}}{np + a\sqrt{np}}\mathbb{P}_{\mathcal{C}}(\mathcal{D}) + \frac{3n}{2} - 2 - \frac{3n}{2}\mathbb{E}_{\mathcal{C}}\left[\frac{1}{d(u) + 1}\right]\\ &- \frac{n}{2}\mathbb{E}_{\mathcal{C}}[d(u)^2R(u, i)^2]^{1/2},\end{aligned}$$

which also evaluates to $n(1 - O(\varepsilon n))$. □

5.3 | Proof of Theorems 1.4 and 1.9

Lemma 5.4. *Let $\mathcal{G} \sim \mathcal{G}(,n, p)$ satisfy* (4). *Then* $\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})^2], \mathbb{E}_{\mathcal{C}}[T(\mathcal{G})^2] = n^2(1 \pm O\ (\varepsilon_n))$.

Proof. We will first bound $\mathbb{E}_{\mathcal{C}}[h(i,j)^3]$ from above. Now similarly to Lemma 5.3,

$$\mathbb{E}_{\mathcal{C}}[h(i,j)^3] = \frac{1}{8}\sum_{x,y,z\in V}\mathbb{E}_{\mathcal{C}}[d(x)d(y)d(z)R(i,j)^3] = \frac{1}{8}\sum_{x,y,z\in V}(8 \pm O(\epsilon_n)), \tag{63}$$

which equals $n^3(1 \pm O\ (\varepsilon_n))$—where above we applied Tetali's formula (3), cancelled terms by symmetry and then applied Lemma 5.2 with $A = \{x, y, z\}$ and the multi-set $B = \{(i,j), (i,j), (i,j)\}$.

By the definition (5) of $T(\mathcal{G})$ and Hölder's inequality (A2) we have

$$\begin{aligned}\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})^2] = \mathbb{E}_{\mathcal{C}}\left[\left(\sum_{j\in V}\frac{d(j)}{2m}h(i,j)\right)^2\right] &= \mathbb{E}_{\mathcal{C}}\left[\sum_{j,k\in V}\frac{d(j)d(k)}{(2m)^2}h(i,j)h(x,y)\right]\\ &\leq \sum_{j,k\in V}(\mathbb{E}_{\mathcal{C}}[d(j)^9]\mathbb{E}_{\mathcal{C}}[d(k)^9]\mathbb{E}_{\mathcal{C}}[1/(2m)^{18}])^{1/9}(\mathbb{E}_{\mathcal{C}}[h(i,j)^3]\mathbb{E}_{\mathcal{C}}[h(x,y)^3])^{1/3}.\end{aligned}$$

Applying the bounds (A4), (62), and (63) respectively then Bernoulli's inequality (A1) gives

$$\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})^2] \leq \frac{n^2}{2^2}((np)^9 + O\left((np)^8\right))^{\frac{2}{9}}\left(\frac{2^{18} + O(\varepsilon_n)}{n^{36}p^{18}}\right)^{\frac{1}{9}}(n^6\ (1 + O(\varepsilon n)))^{\frac{1}{3}} = n^2(1 + O(\varepsilon n)).$$

Then by Jensen's inequality and the lower bound on $\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})]$ proved earlier we have

$$\mathbb{E}_{\mathcal{C}}[T(\mathcal{G})^2] \geq \mathbb{E}_{\mathcal{C}}[T(\mathcal{G})]^2 \geq (n\left(1 - O(\varepsilon n)\right))^2 = n^2\left(1 - O(\varepsilon n)\right).$$

The exact same calculations yield the same bounds for $\mathbb{E}_{\mathcal{C}}[H_i(\mathcal{G})^2]$. □

We prove Theorems 1.4 and 1.9 (together) by Chebychev's inequality and our moment bounds.

Proof of Theorems 1.4 *and* 1.9. Let $X \in \{h(i,j),\ H_i(\mathcal{G}),\ T(\mathcal{G}),\ cc_i\}$ where $i,j \in V$ and recall $\mathbb{E}_{\mathcal{C}}[\cdot] = \mathbb{E}[\cdot|\mathcal{C}]$. We have the following for these X by Theorem 1.2

$$\mathrm{Var}(X\,|\,\mathcal{C}) = n^2(1 + O(\varepsilon n)) - (n(1 + O(\varepsilon n)))^2 = O\left(n^2\varepsilon_n\right).$$

We can also calculate the conditional variance of $K(\mathcal{G})$ by Theorem 1.2, this yields

$$\mathrm{Var}(K(\mathcal{G})|\mathcal{C}) = (n^2/p^2)(1 + O(\varepsilon n)) - (n(1 + O(\varepsilon n))/p)^2 = O(n\varepsilon_n/p).$$

By the Chebyshev inequality [2, Theorem 4.1.1] for each of the above

$$\mathbb{P}(|X - \mathbb{E}[X\,|\,\mathcal{C}]| \geq \lambda(n)\sqrt{\mathrm{Var}(X\,|\,\mathcal{C})}\,|\mathcal{C}) \leq 1/\lambda(n)^2.$$

For X above we have $\mathrm{Var}(X\,|\,\mathcal{C}) = O\left(\mathbb{E}[X\,|\,\mathcal{C}]^2\varepsilon_n\right)$ by Theorem 1.2, thus there exists some K independent of n and X such that $\sqrt{\mathrm{Var}(X\,|\,\mathcal{C})} < \mathbb{E}[X\,|\,\mathcal{C}]\sqrt{K\varepsilon_n}$, for large n. By choosing $\lambda(n) = \sqrt{f(n)/K}$ for any function $f(n)$ we have

$$\mathbb{P}(|X - \mathbb{E}[X\,|\,\mathcal{C}]| > \mathbb{E}[X\,|\,\mathcal{C}]\sqrt{f(n)\varepsilon_n}\,|\mathcal{C}) \leq K/f(n) = O(1/f(n)).$$

The result follows since $\mathbb{P}(\mathcal{E}) \leq \mathbb{P}(\mathcal{E}|\mathcal{C}) + \mathbb{P}(\mathcal{C}^c)$, for any event $\mathcal{E}$. □

ACKNOWLEDGMENTS

I would like to thank my supervisors Agelos Georgakopoulos and David Croydon for their guidance and the reviewers who's comments have greatly improved the paper. This work was started while I was part of the MASDOC DTC at the University of Warwick, supported by EPSRC grant no. EP/HO23364/1 and ERC starting grant no. 639046 (RGGC). It was completed while I supported by ERC starting grant no. 679660 (DYNAMIC MARCH).

ORCID

John Sylvester http://orcid.org/0000-0002-6543-2934

REFERENCES

1. D. Aldous and J. A. Fill, *Reversible Markov chains and random walks on graphs*, 2002. Unfinished monograph, recompiled 2014.
2. N. Alon, and J. Spencer, *The probabilistic method*, Wiley, New York, NY.
3. B. Bollobás, Modern graph theory, vol. **184**, Springer Science & Business Media, New York, NY, 1998.
4. B. Bollobás, *Random graphs*, Cambridge Studies in A Mathematics, 2nd ed., Cambridge University Press, Cambridge, 2001, p. xviii+498.
5. N. Boumal and X. Cheng, *Concentration of the Kirchhoff index for Erdős-Rényi graphs*, Systems Control Lett. **74** (2014), 74–80.
6. A. Z. Broder, A. M. Frieze, S. Suen, and E. Upfal, *Optimal construction of edge-disjoint paths in random graphs*, SIAM J. Comput. **28** (1999), no. 2, 541–573.
7. A. K. Chandra, P. Raghavan, W. L. Ruzzo, R. Smolensky, and P. Tiwari, *The electrical resistance of a graph captures its commute and cover times*, Comput. Complex. **6** (1997), no. 4, 312–340.
8. F. Chung and L. Lu, *The diameter of sparse random graphs*, Adv. Appl. Math. **26** (2001), no. 4, 257–279.
9. F. Chung and L. Lu. Complex graphs and networks, volume 107 of *CBMS Regional Conference Series in Mathematics*. 2006.
10. A. Coja-Oghlan, *On the Laplacian eigenvalues of* $G_{n,p}$, Combin. Probab. Comput. **16** (2007), no. 6, 923–946.
11. C. Cooper and A. Frieze, *The cover time of sparse random graphs*, Random Struct. Algorithms **30** (2007), no. 1-2, 1–16.
12. M. Randić and D. J. Klein, *Resistance distance*, J. Math. Chem. **12** (1993), no. 1, 81–95.
13. P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, Mathematical Association of America, Washington, DC, 1984.
14. L. Erdős, A. Knowles, H. -T. Yau, and J. Yin, *Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues*, Comm. Math. Phys. **314** (2012), no. 3, 587–640. https://doi.org/10.1007/s00220-012-1527-7
15. P. Erdős and A. Rényi, *On random graphs. I*, Publ. Math. Debrecen. **6** (1959), 290–297.
16. A. Frieze, and M. Karoński, Introduction to random graphs, Cambridge University Press, Cambridge, 2015.
17. A. Georgakopoulos, *A tractable variant of cover time*, arXiv. **1206** (2012), 6605.
18. A. Georgakopoulos and S. Wagner, Hitting times, cover cost, and the wiener index of a tree. J. Graph Theory **84** (2017), no. 3, 311–326.
19. S. Janson, T. Łuczak, and A. Rucinski, *Random graphs*Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, NY, 2000.
20. J. Jonasson, *On the cover time for random walks on random graphs*, Combin. Probab. Comput. **7** (1998), no. 3, 265–279.
21. A. Knoblauch, *Closed-form expressions for the moments of the binomial probability distribution*, SIAM J. Appl. Math. **69** (2008), no. 1, 197–204.
22. D. A. Levin, Y. Peres, and E. L. Wilmer, Markov chains and mixing times, American Mathematical Society, Providence, RI, 2009.
23. L. Lovász. *Random walks on graphs: A survey*. In Combinatorics, Paul Erdős is eighty, vol. 2 (Keszthely, 1993), volume 2 of Bolyai Soc. Math. Stud.,pp. 353–397. János Bolyai Math. Soc., Budapest, 1996.
24. M. Löwe and F. Torres, *On hitting times for a simple random walk on dense Erdös-Rényi random graphs*, Stat. Probab. Lett. **89** (2014), 81–88.
25. R. Lyons, and Y. Peres, Probability on trees and networks, Cambridge University Press, New York, NY, 2016.
26. P. Tetali, *Random walks and the effective resistance of networks*, J. Theoret. Probab. **4** (1991), no. 1, 101–109.
27. U. vonLuxburg, A. Radl, and M. Hein, *Hitting and commute times in large random neighborhood graphs*, J. Mach. Learn. Res. **15** (2014), 1751–1798.

28. D. B. Wilson. Generating random spanning trees more quickly than the cover time. In *Proceedings of the Twenty-eighth Annual ACM Symposium on the Theory of Computing* (Philadelphia, PA, 1996), ACM, New York, NY, 1996, pp. 296–303.

How to cite this article: Sylvester J. Random walk hitting times and effective resistance in sparsely connected Erdős-Rényi random graphs. *J Graph Theory*. 2020;1–41. https://doi.org/10.1002/jgt.22551

APPENDIX A

We make frequent use of the following inequalities. Bernoulli: Let $x \geq -1$, then

$$(1+x)^r \leq 1 + rx \text{ for } 0 \leq r \leq 1 \quad \text{and} \quad (1+x)^r \geq 1 + rx \text{ for } r \geq 1. \tag{A1}$$

Hölder: For $1 \leq k \leq n$ let X_k be r.v.'s and $p_k \in [1, \infty)$, where $\sum_{k=1}^{n} 1/p_k = 1$ and $\mathbb{E}[X_k^{p_k}]$ exists, then

$$\mathbb{E}[X_1 \cdots X_n] \leq \mathbb{E}[X_1^{p_1}]^{1/p_1} \cdots \mathbb{E}[X_n^{p_n}]^{1/p_n}. \tag{A2}$$

Coupling: If X, Y are real random variables on a probability space $(\Omega, \mathfrak{F}, \mathbb{P})$, then for any $B \subset \mathbb{R}$,

$$|\mathbb{P}(X \in B) - \mathbb{P}(Y \in B)| \leq \mathbb{P}(X \neq Y). \tag{A3}$$

Lemma A.1 (Chernoff bounds). *If $X \sim Bin(n, p)$, then for any $a > 0$, $b < np$, and $c > np$*

(i) $\mathbb{P}[X < np - a] \leq \exp(-\frac{a^2}{2np})$, and $\mathbb{P}[X > np + a] \leq \exp(-\frac{a^2}{2(np + a/3)})$,
(ii) $\mathbb{P}[X < b] \leq e^{-np}(\frac{enp}{b})^b$, and $\mathbb{P}[X > c] \leq e^{-np}(\frac{enp}{c})^c$.

Proof. For (i) see [9, Theorem 2.4] and [9, Theorem 2.15] with $\varepsilon = 1 - \frac{b}{np}$ for (ii). $\square$

We also have the following closed form for moments of binomial random variables,

Theorem A.2 (Knoblauch [21, Theorem 4.1]). *Let $X \sim Bin(n, p)$, $n^{\underline{i}} := n(n-1)\ldots(n-i+1)$ and $S(d, i)$ be the Stirling partition number of d items into i subsets. Then for $d \geq 0$,*

$$\mathbb{E}[X^d] = \sum_{i=0}^{d} S(d, i) p^i n^{\underline{i}}, \quad \text{where} \quad S(d, i) := \frac{1}{i!} \sum_{k=0}^{i} (-1)^{k+i} \binom{i}{k} k^d.$$

Let $X \sim Bin(n, p)$, $0 < p := p(n) < 1$ and $d \geq 0$ fixed. Then by Theorem A.2 we have

$$\mathbb{E}[X^d] = S(d, d) p^d n^{\underline{d}} \pm O(p^{d-1} n^{\underline{d-1}}) = (np)^d \pm O((np)^{d-1}). \tag{A4}$$

Proposition A.3. *Let $X \sim Bin(n, p)$, $Y \sim Bin(n-1, p)$, $\alpha \in \mathbb{Z}, \alpha \geq 1$. Then*

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{X\geq 1\}}}{X^{\alpha}}\right] := \sum_{k=1}^{n}\frac{1}{k^{\alpha}}\binom{n}{k}p^{k}(1-p)^{n-k} = \sum_{k=0}^{n-1}\frac{1}{(k+1)^{\alpha}}\binom{n}{k+1}p^{k+1}(1-p)^{n-1-k}$$
$$= \sum_{k=0}^{n-1}\frac{np}{(k+1)^{\alpha+1}}\binom{n-1}{k}p^{k}(1-p)^{(n-1)-k} = \mathbb{E}\left[\frac{np}{(Y+1)^{\alpha+1}}\right].$$

Lemma A.3. *Let $X_n \sim Bin(n,p)$ for $p := p(n)$ with $np \to \infty$, $a \in \mathbb{R}$, $b \in \mathbb{Z}$, $a, b > 0$. Then*

$$\frac{1}{(a+np)^{b}} \leq \mathbb{E}\left[\frac{1}{(a+X_n)^{b}}\right] \leq \frac{1}{(a+np)^{b}} + O\left(\frac{1}{(np)^{(b+1)}}\right).$$

Proof. Let $f(x) := f_{a,b}(x) = (a+x)^{-b}$ for any constants $a, b > 0$. The lower bound follows from Jensen's inequality since $f(x)$ is convex for $a, b > 0$.

Let $\mu_n = \mathbb{E}[X_n] = np$. When $np \to \infty$ it is possible to find some $r := r(n)$ such that $r = \omega(\sqrt{np}\log(np))$ and $r = o(np)$. The Chernoff bound, Lemma A.1 (i), then yields

$$\mathbb{P}(X_n \leq \mu_n - r) \leq \exp(-r^2/2\mu_n) = o(1/np).$$

With this r we can achieve the following a priori upper bound for any $b \geq 1$:

$$\mathbb{E}[f(X_n)] \leq \frac{1}{a^{b}}\mathbb{P}(X_n \leq \mu_n - r) + f(\mu_n - r)\mathbb{P}(X_n > \mu_n - r) = (1+o(1))f(\mu_n). \tag{A5}$$

By Taylor's theorem there is some ξ_n between X_n and μ_n such that

$$f(X_n) = f(\mu_n) + f'(\mu_n)(X_n - \mu_n) + f''(\xi_n)(X_n - \mu_n)^2.$$

Using Hölder's inequality (A2) and the fact $f(x)$ is decreasing when $x > 0$, we have

$$\begin{aligned}(\mathbb{E}[f(X_n)] - f(\mu_n))^2 &\leq (f'(\mu_n)\mathbb{E}[X_n - \mu_n] + \mathbb{E}[f''(\xi_n)(X_n - \mu_n)^2])^2\\ &\leq \mathbb{E}[f''(\xi_n)^2]\mathbb{E}[(X_n - \mu_n)^4] \leq \mathbb{E}[f''(X_n)^2\mathbf{1}_{\{X_n \leq \mu_n\}}]\mathbb{E}[(X_n - \mu_n)^4]\\ &\quad + \mathbb{E}[f''(\mu_n)^2\mathbf{1}_{\{X_n > \mu_n\}}]\mathbb{E}[(X_n - \mu_n)^4] \leq (2+o(1))f''(\mu_n)^2\mathbb{E}[(X_n - \mu_n)^4].\end{aligned} \tag{A6}$$

The last inequality follows by (A5) since $f''(\mu_n) = b\cdot(b+1)\cdot(a+\mu_n)^{-(b+2)}$. Observe

$$\mathbb{E}[(X_n - \mu_n)^4] = np(1-p)(3p(n-2) - 3p^2(n+2) + 1) = O((np)^2), \tag{A7}$$

this can be calculated using the binomial moment generating function or by Theorem A.2. Hence by (A6), (A7), and $(f_{a,b}(x))'' = b(b+1)f_{a,(b+2)}(x)$, we have

$$\mathbb{E}[f(X_n)] \leq f(\mu_n) + \left(O\left((a+\mu_n)^{-2(b+2)}\right)\cdot O((np)^2)\right)^{1/2} = \frac{1}{(a+np)^{b}} + O\left(\frac{1}{(np)^{b+1}}\right). \quad \square$$

Finally we shall prove Proposition 1.5 which shows tightness for the concentration results.

Proof of Proposition 1.5. Let X_d be the number of vertices with degree d. For the first case:

$$\mathbb{E}[X_1] = n\cdot\binom{n-1}{1}p(1-p)^{n-2} = n^2pe^{-\log n - O(\log\log\log n)} \geq \frac{\log n}{(\log\log n)^{O(1)}}.$$

This implies that, for any fixed t, $\lim_{n\to\infty}\mathbb{P}(|X_1| \geq t) = 1$ by [4, Theorem 3.1]. Thus w.h.p. there is at least one pair of vertices i, j both with degree 1 and so $R(i, j) \geq 1$ by Lemma 2.1. Since the number of edges m is distributed $Bin(\binom{n}{2}, p)$ there are $n^2p/2(1 - o(1))$ edges w.h.p by Lemma A.1. Thus by the Commute time formula (9) we have $\kappa(i, j) = 2m\cdot R(i, j) \geq (1 - o(1))n\log(n)$ and since $\kappa(i, j) = h(i, j) + h(j, i)$ at least one of $h(i, j)$ of $h(j, i)$ is greater than $n\log(n)/3$ w.h.p.

For the case $np = (c + o(1))\log(n)$ let $k : k(\varepsilon) := (1 - \varepsilon)np$ for some $0 < \varepsilon < 1$ and observe

$$\mathbb{E}[X_k] = n\binom{n-1}{k}p^k(1-p)^{n-1-k} \geq \frac{n}{\sqrt{2\pi k}}\left(\frac{e}{(1-\varepsilon)}\right)^{(1-\varepsilon)np}e^{-np}(1 - o(1)).$$

Recall $-\log(1 - t) \geq t + t^2/2$ for $t < 1$. In a similar vein to the proof of [20, Theorem 2.2]:

$$\mathbb{E}[X_k] \geq \frac{n}{3\sqrt{k}}e^{-\varepsilon np-(1-\varepsilon)\log(1-\varepsilon)np} \geq \frac{n}{3\sqrt{k}}e^{-\varepsilon np+(1-\varepsilon)(+\varepsilon+\varepsilon^2/2)np} \geq \frac{ne^{-\frac{\varepsilon^2(1+\varepsilon)np}{2}}}{3\sqrt{k}}.$$

So for any $0 < \varepsilon < 1$ satisfying $\frac{\varepsilon^2(1+\varepsilon)c}{2} < 1$ we have that $\mathbb{E}[X_k] \to \infty$ (a concrete example would be $\varepsilon = \sqrt{1/(c + 1)}$). Thus again by [4, Theorem 3.1] there are at least two vertices i, j with degree less than $(1 - \varepsilon)np$ w.h.p. Thus, as before, $\kappa(i, j) = 2m\cdot R(i, j) \geq n^2p\frac{2}{(1-\varepsilon)np} = \frac{2n}{1-\varepsilon}$. Thus one or both of $h(i, j)$ or $h(j, i)$ must be greater than $(1 + a)n$ for some $a > \frac{\varepsilon}{2(1-\varepsilon)} > 0$ w.h.p. □